



République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique



Université de Batna 2 – Mostefa Ben Boulaid -

Faculté/institut de : Lettres et Langues Étrangères  
Département de : Français

## THESE

Présentée pour l'obtention du titre de :

Docteur en français

Option : didactique des langues  
étrangères

Sous le thème :

**La validité de l'évaluation sommative du FLE au  
cycle moyen en Algérie : de l'étude à la conception  
d'un dispositif d'évaluation**

Présentée par :

**LAIDOUDI Assia**

Devant le jury composé de :

- |                              |            |                       |           |
|------------------------------|------------|-----------------------|-----------|
| • M. DAKHIA Abdelouahab      | Professeur | Université de Biskra  | Président |
| • Mme KHADRAOUI Errime       | MCA        | Université de Batna 2 | Encadreur |
| • M. BELKACEM Mohammed Amine | MCA        | Université de Batna 2 | Examineur |
| • Mme AISSI Radhia           | MCA        | Université de Batna 2 | Examineur |
| • M. SLITANE Kamel           | MCA        | Université de M'sila  | Examineur |
| • Mme AMRANI Amira Khadoudja | MCA        | Université de Guelma  | Examineur |

2020/2021

## REMERCIEMENTS

*Je tiens à exprimer mes sincères remerciements à ma directrice de recherche, Mme KHADRAOUI Errime. Je la remercie de m'avoir encadrée, orientée et conseillée.*

*Mes remerciements sont également destinés :*

*Aux membres du jury qui ont accepté de lire et d'évaluer mon travail,*

*Aux inspecteurs qui ont évalué le dispositif proposé,*

*A ma famille,*

*A mes amis qui m'ont toujours soutenue*

*Et à tous ceux qui liront un jour ce travail.*

## Table des matières

<i>Liste des tableaux</i> .....	7
<i>Liste des figures</i> .....	8
<i>Liste des graphiques</i> .....	9
<b>INTRODUCTION GENERALE</b> .....	<b>12</b>
<b>PARTIE THÉORIQUE</b> .....	<b>22</b>
<b>CHAPITRE 1 : AUTOUR DE L'ÉVALUATION SOMMATIVE</b> .....	<b>23</b>
<b>Introduction</b> .....	<b>24</b>
1. A propos de l'évaluation .....	25
1.1. Deux paradigmes en évaluation.....	26
1.1.1. Le paradigme ancien (le paradigme du test theory) .....	26
1.1.2. Le paradigme rénové (Évaluation liée à l'apprentissage).....	28
1.2. Vers une définition de l'évaluation.....	31
1.2.1. Évaluation : un jugement de valeur.....	31
1.2.2. Évaluation : collecte des informations .....	32
1.2.3. Évaluation : notation et prise de décision.....	33
1.2.4. Évaluation : un processus de régulation .....	34
1.3. Types d'évaluation .....	35
1.3.1. Évaluation diagnostique (prédictive) .....	37
1.3.2. Évaluation formative (régulatrice).....	38
1.3.3. Évaluation sommative.....	39
1.4. Les évaluateurs dans le contexte éducatif .....	40
1.4.1. Évaluateurs distincts des formateurs.....	40
1.4.2. Enseignants-évaluateurs.....	41
1.4.3. Apprenants-évaluateurs.....	41
1.5. Fonctions de l'évaluation .....	43
1.6. Qu'évalue-t-on dans le milieu scolaire ?.....	43
2. Évaluation sommative : définitions .....	44
2.1. Évaluer au terme des apprentissages.....	45
2.2. Buts particuliers à l'évaluation sommative .....	46
2.3. Un jugement définitif.....	48
2.4. Un cadre de référence double .....	49
2.5. Notation : exclusive à l'évaluation sommative ? .....	50
3. Croisement des formes : réalité ou mythe ? .....	52
3.1. Évaluation sommative : pourrait-elle devenir formative ? .....	52
3.2. Une distinction indispensable de la forme certificative.....	53
4. Professionnalisme de l'évaluation sommative : quels critères ? .....	54
4.1. La validité .....	55
4.2. La fidélité (fiabilité).....	56
4.3. L'objectivité .....	58
4.4. La faisabilité.....	58
4.5. L'équité.....	59
4.6. La validité : pour une évaluation authentique.....	60
5. Fonctions de l'évaluation sommative.....	62
6. Évaluation sommative : une forme souvent contestée .....	62
7. Des pratiques qui relèvent de l'évaluation sommative .....	66
7.1. Les examens.....	67
7.1.1. Les épreuves écrites .....	70
7.1.2. Les rédactions.....	71
7.1.3. Les examens oraux .....	72
7.2. Les tests .....	74
7.3. Pour une évaluation scolaire : tests ou examens.....	75
<b>Conclusion</b> .....	<b>76</b>
<b>CHAPITRE 2 : LA VALIDITE ET LA VALIDATION DES EPREUVES</b> .....	<b>79</b>
<b>Introduction</b> .....	<b>80</b>
1. La validité : des conceptions définitoires.....	82

1.1 La validité : une qualité du test.....	82
1.2. La validité : une caractéristique de l'interprétation des résultats .....	85
2. Les formes (types) de la validité.....	88
2.1. La validité du contenu.....	89
2.2. La validité prédictive .....	90
2.3. La validité du construit .....	91
2.4. La validité écologique.....	93
2.5. La validité déontologique.....	93
2.6. La validité concourante.....	94
2.7. La validité incrémentale.....	94
2.8. Et la fiabilité ?.....	94
3. Des conceptions de validité aux procédures de validation .....	95
3.1. Des formes de validité aux procédures de validation.....	96
3.1.1. Pour la validité prédictive .....	96
3.1.2. Pour la validité du contenu.....	96
3.1.3. Pour la validité du construit.....	97
3.1.4. Pour la validité concourante .....	97
3.2. Les modèles de validation .....	98
3.2.1. Les travaux de Messick (1994).....	98
3.2.2. Les travaux de Toulmin (1958) .....	99
3.2.3. The Evidence-Centered Design (2003).....	100
3.2.4. Structuration de l'argumentation interprétative (Kane, 2006) .....	103
3.2.4.1. La notation .....	104
3.2.4.2. La généralisation .....	109
3.2.4.3. L'extrapolation .....	110
3.2.4.4. L'implication.....	112
3.2.4.5. Limites du modèle de Kane (2006).....	114
3.2.5. Validation des épreuves selon De Ketele et Gérard.....	115
3.2.5.1. La validation a priori par recours à des juges .....	115
3.2.5.2. Les validations a posteriori .....	117
4. Les biais de l'évaluation sommative .....	121
4.1. La planification .....	121
4.2. La conception des épreuves.....	123
4.3. La correction des copies.....	125
<b>Conclusion .....</b>	<b>129</b>
<b>CHAPITRE 3 : DES ETAPES POUR LA CONCEPTION DES DISPOSITIFS EVALUATIFS .....</b>	<b>131</b>
<b>Introduction .....</b>	<b>132</b>
1. Qu'est-ce qu'un dispositif d'évaluation ?.....	133
2. La planification : une étape-clé vers un dispositif valide.....	135
2.1. Que faut-il évaluer dans le cadre de l'APC ? .....	136
2.1.1. La compétence : des définitions .....	136
2.1.2. Peut-on évaluer les compétences ? .....	140
2.1.3. Peut-on évaluer les ressources dans toute évaluation scolaire ? .....	144
2.2. Des précisions sur les examens sommatifs à concevoir .....	145
2.2.1. Le moment, nombre et durée des examens.....	146
2.2.2. Norme à fixer .....	147
2.2.3. Communication des résultats : quels outils ?.....	147
2.2.4. Conditions de passation à établir .....	148
3. Conception des instruments de mesure et leur expérimentation .....	148
3.1. Construction des outils de mesure pour une classe de FLE.....	149
3.1.1. Que faut-il évaluer ?.....	149
3.1.2. Évaluer par des situations.....	151
3.2. La situation d'évaluation.....	151
3.2.1. Caractéristiques des situations d'évaluation .....	151
3.2.2. Composantes des situations d'évaluation complètes.....	158
3.2.2.1. Les supports .....	158

3.2.2.2. Les tâches.....	159
3.2.2.3. Les consignes .....	162
3.3. Conception de situations évaluatives : quelles étapes ? .....	165
3.3.1. Identifier la compétence évaluée et les caractéristiques mobilisées .....	166
3.3.2. Élaborer une situation conforme aux exigences .....	166
3.3.3. Déterminer les critères de correction .....	167
3.3.4. Vérifier si la situation correspond aux exigences .....	168
3.3.5. Expérimentation .....	168
4. Passation des épreuves .....	169
5. Jugement : quels outils faut-il intégrer ? .....	170
5.1. La liste de vérification .....	170
5.2. Le barème de notation.....	171
5.3. La grille d'évaluation.....	171
5.3.1. Qu'est-ce qu'une grille ?.....	172
5.3.2. Pourquoi le recours aux grilles d'évaluation ?.....	173
5.3.3. Conception des grilles : quelles étapes faut-il suivre ?.....	174
5.3.3.1. Le choix des critères d'évaluation .....	175
5.3.3.2. La sélection de l'échelle d'appréciation .....	177
5.3.3.3. Déterminer un moyen pour exprimer le jugement .....	180
5.3.3.4. Assembler la grille.....	181
5.3.3.5. Expérimenter la grille conçue.....	182
6. La décision .....	182
<b>Conclusion .....</b>	<b>183</b>
<b><i>PARTIE PRATIQUE</i>.....</b>	<b>186</b>
<b><i>CHAPITRE 1 : CADRE METHODOLOGIQUE</i>.....</b>	<b>187</b>
<b>Introduction .....</b>	<b>188</b>
1. Méthodologie .....	189
1.1. L'analyse qualitative et quantitative (approche mixte) .....	190
1.2. La méthode comparative .....	191
2. Les instruments de mesure .....	192
2.1. La grille d'analyse .....	193
2.1.1. Les précisions sur l'épreuve .....	193
2.1.2. Les compétences évaluées.....	195
2.1.3. Les supports .....	196
2.1.4. Les outils de mesure .....	199
2.1.5. Les outils de jugement .....	209
2.2. La liste de vérification .....	213
3. Contexte de l'étude.....	215
3.1. A propos des types d'évaluation.....	216
3.2. Quels outils évaluatifs ?.....	219
4. Des précisions sur les corpus analysés.....	226
4.1. Corpus 1 : devoirs et compositions.....	227
4.1.1. Taille du corpus.....	227
4.1.2. Bornes chronologiques.....	229
4.1.3. Bornes géographiques.....	229
4.1.4. Méthodes de collecte .....	230
4.2. Corpus 2 : les examens de BEM.....	231
4.2.1. Taille.....	232
4.2.2. Précisions chronologiques.....	232
4.2.3. Méthodes de collecte .....	233
<b>Conclusion .....</b>	<b>234</b>
<b><i>CHAPITRE 2 : RÉSULTATS, ANALYSE ET DISCUSSION</i>.....</b>	<b>236</b>
<b>Introduction .....</b>	<b>237</b>
1. A propos de notre corpus .....	238
2. Résultats et analyse.....	241
2.1. Concernant les compétences évaluées .....	241

2.2. Les supports : quelle conformité aux instructions ?.....	243
2.3. Les outils de mesure .....	254
2.4. Les outils de jugement .....	277
3. Discussion .....	279
3.1. Les compétences évaluées.....	280
3.2. Les supports sélectionnés.....	282
3.3. La compréhension et les ressources : outils intégrés.....	285
3.4. La production écrite : comparaison des situations .....	287
3.5. Les outils de jugement .....	288
<b>Conclusion .....</b>	<b>289</b>
<b>CHAPITRE 3 : PROPOSITION DE DISPOSITIF ÉVALUATIF VALIDE .....</b>	<b>292</b>
<b>Introduction .....</b>	<b>293</b>
1. Les compétences à évaluer au cycle moyen .....	294
2. Des propositions de situations évaluatives .....	296
3. Vérification des situations conçues.....	297
4. Expérimentation .....	299
5. Proposition de grilles de correction .....	302
6. Dispositif évaluatif proposé.....	306
6.1. Première année moyenne .....	306
6.2. Deuxième année moyenne .....	313
6.3. Troisième année moyenne.....	318
6.4. Quatrième année moyenne.....	324
<b>Conclusion .....</b>	<b>329</b>
<b>CONCLUSION GENERALE.....</b>	<b>332</b>
<b>BIBLIOGRAPHIE.....</b>	<b>339</b>
<b>ANNEXES .....</b>	<b>362</b>

# Liste des tableaux

<b>Tableau 1: Types d'évaluation.....</b>	<b>36</b>
<b>Tableau 2:Exemple d'une situation d'intégration .....</b>	<b>220</b>
<b>Tableau 3:Grille de l'enseignant .....</b>	<b>222</b>
<b>Tableau 4: Exemple d'une épreuve d'évaluation.....</b>	<b>227</b>
<b>Tableau 5: Grille d'évaluation de l'épreuve certificative.....</b>	<b>228</b>

# Liste des figures

<b>Figure 1: Structuration argumentative de Kane.....</b>	<b>104</b>
<b>Figure 2:Schéma des étapes de la conception de dispositifs valides .....</b>	<b>135</b>
<b>Figure 3:les types des échelles d'appréciation.....</b>	<b>178</b>



# Liste des graphiques

Graphique 1: Répartition des types d'évaluations constituant notre corpus .....	239
Graphique 2: Répartition des niveaux du cycle moyen introduits dans l'étude .....	239
Graphique 3: Répartition des trimestres scolaires couverts par notre corpus .....	239
Graphique 4: Répartition des années scolaires englobées par notre corpus .....	240
Graphique 5: Durée des épreuves constituant notre corpus .....	240
Graphique 6: Répartition des compétences évaluées .....	241
Graphique 7: Répartition de la manière évaluant les deux compétences .....	242
Graphique 8: Outils de mesure évaluant les différentes compétences .....	242
Graphique 9: Nombre des supports dans les épreuves des différents niveaux .....	243
Graphique 10: Les parties des épreuves qui contiennent des supports .....	244
Graphique 11: Nature des supports proposés dans notre corpus .....	244
Graphique 12: Thèmes des deux parties (Compréhension et Production) .....	245
Graphique 13: Les supports des épreuves : entre inédits et déjà exploités .....	247
Graphique 14: Les supports de notre corpus : authentiques, adaptés ou fabriqués ? .....	248
Graphique 15: Les sources dans les textes proposés dans les épreuves .....	250
Graphique 16: Longueur des textes proposés en 1AM .....	251
Graphique 17: Longueur des textes proposés en 2AM .....	251
Graphique 18: Longueur des textes proposés en 3AM .....	252
Graphique 19: Longueur des textes proposés en 4AM .....	253
Graphique 20: Nombre de questions évaluant la compréhension dans les sujets analysés .....	254
Graphique 21: Nombre des questions évaluant les ressources dans les sujets analysés .....	255
Graphique 22: Comparaison du nombre des questions des deux catégories (Compréhension et Ressources) .....	255
Graphique 23: Représentation des types d'activités évaluant la compréhension .....	256
Graphique 24: Répartition des activités évaluant les ressources .....	259
Graphique 25: Les niveaux de compréhension évalués dans les épreuves analysées .....	262
Graphique 26: Les niveaux cognitifs des items évaluant les ressources .....	264
Graphique 27: Questions interdépendantes vs questions indépendantes .....	265
Graphique 28: Types évaluant la compétence expressive dans les épreuves analysées .....	267
Graphique 29: Les situations inédites dans le corpus analysé .....	269
Graphique 30: La précision du contexte dans les situations évaluatives .....	271
Graphique 31: Répartition des précisions dans les contextes des situations .....	272
Graphique 32: Les constituants des situations analysées .....	273
Graphique 33: L'utilité des supports intégrés dans les situations évaluatives .....	273
Graphique 34: Les informations dans les situations d'évaluation du cycle moyen .....	274
Graphique 35: Le respect de la règle des 2/3 dans les situations du corpus .....	275
Graphique 36: Les instructions dans les situations d'évaluation .....	276
Graphique 37: La nature des instructions que les situations contiennent .....	276
Graphique 38: Les outils de jugement investis dans les épreuves du FLE .....	277
Graphique 39: Les parties évaluées par le barème de notation .....	278

<b>Graphique 40: Le taux accordé aux ressources dans les épreuves</b> .....	279
<b>Graphique 41: Répartition des compétences évaluées dans les épreuves</b> .....	281
<b>Graphique 42: Compétences évaluées séparément ou en parallèle</b> .....	281
<b>Graphique 43: Outils évaluant la compréhension</b> .....	281
<b>Graphique 44: Outils évaluant les ressources</b> .....	282
<b>Graphique 45: Outils évaluant la production écrite</b> .....	282
<b>Graphique 46: Nombre de supports sélectionnés dans les BEM analysés</b> .....	283
<b>Graphique 47: Parties qui intègrent des supports</b> .....	283
<b>Graphique 48: Nature des supports investis</b> .....	283
<b>Graphique 49: Supports des épreuves analysées : authentiques, adaptés ou fabriqués ?</b> .....	283
<b>Graphique 50: Précisions sur les sources investies</b> .....	284
<b>Graphique 51: Longueur des supports sélectionnés dans les BEM</b> .....	284
<b>Graphique 52: Comparaison des questions évaluant la compréhension à celles évaluant les ressources</b> .....	285
<b>Graphique 53: Types de questions évaluant la compréhension dans les BEM</b> .....	285
<b>Graphique 54: Niveaux de compréhension évalués</b> .....	286
<b>Graphique 55: Types des activités évaluant les ressources</b> .....	286
<b>Graphique 56: Les niveaux cognitifs des questions qui évaluent les ressources</b> .....	287
<b>Graphique 57: Les éléments constitutifs des situations évaluatives</b> .....	288

## Liste des abréviations

AM : année moyenne

APC : approche par les compétences

BAC : baccalauréat

BEM : brevet d'enseignement moyen

CECRL : cadre européen commun de références pour les langues

ECD : evidence-certred design

FLE : français langue étrangère

PPO : pédagogie par objectifs

# **INTRODUCTION GENERALE**

L'évaluation en tant que concept et notamment en tant que processus est devenue « *au cours de la décennie 1990 l'une des préoccupations majeures de la didactique des langues étrangères et en particulier du FLE* » (Cuq & Gruca, 2005, p. 357), un regain d'intérêt qui s'explique essentiellement par l'évolution du concept et par les contraintes de son application pédagogique. Cette évolution est retracée par de nombreux écrivains à l'instar de Sampsonis et Noël-Jothy (2006) dans leur ouvrage intitulé « *Certification et outils d'évaluation en FLE* ». Ces auteures évoquent le rôle joué par l'évaluation au cours des années 60 : conçue comme un moyen de sanction, elle classe les élèves en catégories. Progressivement, elle devient un processus investi pour la formation complémentaire à l'enseignement/apprentissage. Elle finit par s'imposer comme un « *moyen de guider l'apprentissage* » (Cuq & Gruca, 2005, p. 356). Nous assistons ainsi à de nouveaux liens entre deux processus longtemps dissociés : évaluation et apprentissage sont actuellement des processus complémentaires étant que le premier est intrinsèque au second. Ce sont, en effet, les résultats obtenus en évaluation qui assurent la régulation des pratiques enseignantes et l'ajustement des démarches et des stratégies adoptées.

En dépit de ces liens étroits et complexes entre l'évaluation et l'apprentissage, des voix s'élèvent pour sa suppression du milieu scolaire, une réclamation que certains praticiens sceptiques sur les effets positifs du processus évaluatif admettent aisément (Hadji, 2016). Si cette idée était adoptée, l'apprentissage scolaire serait-il plus efficace ? Nous devons reconnaître que sa suppression -même si elle épargne aux apprenants le stress souvent ressenti et à leurs enseignants certaines peines liées à la conception des examens et à leur correction- aboutit à des apprentissages incertains. En se contentant de transmettre les contenus sélectionnés, les enseignants ne pourraient pas s'assurer de leur assimilation et de leur transfert par leurs apprenants dans les situations scolaires et sociales. Cette suppression requise par certains pourrait révéler leur conception archaïque et restrictive qui réduit l'évaluation à la notation, une conception révolue dans la mesure où les multiples formes évaluatives existantes soutiennent le processus d'apprentissage à tous ses moments.

A travers cette affirmation, nous ne voulons pas rabaisser le statut de la forme sommative de l'évaluation dont les enjeux personnels, professionnels, sociaux...lui assignent une valeur incontestable. Cette forme se définit par son moment d'intégration, son but, sa rétroaction et son cadre de référence : elle intervient au bout des apprentissages, ce qui l'oppose aux formes diagnostique intégrée en amont du processus d'apprentissage et formative introduite au cours de celui-ci ; contrôle les acquis des apprenants pour prendre des décisions ; aboutit à un jugement irrévocable ; et se caractérise par un cadre de référence double (approches

normative et critériée). Mentionnons que l'évaluation dont il est question dans notre étude est celle menée par les enseignants pour évaluer les compétences de leurs apprenants, une précision qui s'impose vu la diversité des évaluateurs, des évalués et des objets évalués.

Toutes ces caractéristiques de l'évaluation sommative seront certainement développées dans le contenu de notre étude. Cependant, celle-ci porte principalement sur les outils à investir pour que les résultats révèlent les acquis effectifs des évalués. L'intérêt que nous accordons à ces outils devient extrême vu la réforme de notre système éducatif algérien où une nouvelle approche (l'Approche Par les Compétences) est introduite rénovant considérablement les pratiques enseignantes et évaluatives. En effet, cette réforme a entraîné la rénovation des programmes, des contenus et des manuels ; ce qui contraint les enseignants à actualiser leurs démarches et leurs pratiques. Vu le lien étroit entre enseignement et évaluation, les pratiques évaluatives se voient dans la nécessité de s'adapter aux exigences de la nouvelle approche ; d'où la nécessité d'introduire de nouveaux outils plus conformes.

Il est constaté, cependant, que les pratiques évaluatives stagnent ; un constat dont la première raison qui s'impose comme une évidence est la stagnation des pratiques enseignantes. En effet, ce qui est évalué est le reflet de ce qui est enseigné et les outils investis pour l'évaluer sont révélateurs des outils investis pour l'enseigner. Si les pratiques évaluatives n'évoluent pas, elles deviennent incapables de donner un aperçu fidèle sur ce qui est réellement appris par les apprenants. Tout écart entre ce qui est visé et ce qui est enseigné puis évalué affecte ainsi la validité des évaluations sommatives ; un concept dont nous nous servons pour désigner la capacité des outils de mesure et de jugement introduits dans les évaluations sommatives du FLE au cycle moyen à mesurer les compétences effectives des apprenants. C'est à cette caractéristique des évaluations sommatives- distinctes des évaluations certificatives où des attestations et des certificats sont attribués aux apprenants pour déterminer ce qui est réellement maîtrisé- que nous nous intéressons dans la présente étude.

Notre travail intitulé « *La validité de l'évaluation sommative du FLE au cycle moyen en Algérie : de l'étude à la conception d'un dispositif d'évaluation* » s'inscrit dans le cadre de la didactique des langues étrangères. Le processus évaluatif est, comme il est antérieurement évoqué, l'un des thèmes complexes de cette didactique que les auteurs continuent à creuser pour la conception d'outils contribuant à améliorer le rendement des apprenants et à fournir un aperçu fidèle sur leurs compétences. Comme les problématiques soulevées par ce thème sont nombreuses, nous nous intéresserons plus particulièrement dans la présente thèse à la validité de l'évaluation sommative. Nous tenterons donc de répondre aux questions suivantes : **à quel**

**point les outils de mesure et de jugement constituant les dispositifs actuels évaluant les compétences des apprenants algériens en FLE au cycle moyen sont-ils valides ? Fournissent-ils des aperçus fidèles sur leurs compétences effectives ? Si ces dispositifs ne sont pas suffisamment valides, quels outils les enseignants peuvent-ils introduire pour une évaluation sommative rigoureuse des compétences visées ?**

Ces questions, auxquelles nous tenterons de répondre tout au long des parties et chapitres constituant notre travail, requièrent des précisions élémentaires répondant à certaines questions connexes ci-dessous citées :

- Quels sont les outils de mesure et de jugement actuellement mis en œuvre par les enseignants du cycle moyen algérien pour évaluer les compétences de leurs apprenants en FLE ?
- Quels sont les outils préconisés dans une perspective de compétences ? Sont-ils introduits actuellement en évaluation sommative ?
- Est-il nécessaire de rénover les pratiques évaluatives du FLE au cycle moyen ? Si oui, cette rénovation serait-elle favorablement accueillie par les enseignants ?
- Les enseignants peuvent-ils évaluer les connaissances en évaluation sommative ?
- La note constitue-t-elle un outil de jugement efficace ? Faut-il la remplacer par des outils plus appropriés ?

A première vue, certaines réponses semblent évidentes à ces questions. En effet, nous supposons que les pratiques évaluatives adoptées par les enseignants du cycle moyen ne sont pas suffisamment valides pour évaluer les compétences effectives de leurs apprenants. Cette hypothèse, nous la formulons en nous appuyant sur deux constats : la ressemblance des épreuves sommatives actuelles du FLE à celles introduites avant la réforme et la prise en compte des connaissances en évaluation. Il paraît en effet que ces épreuves, qui semblent identiques sur le plan structurel aux anciennes, maintiennent les mêmes pratiques et que les ressources linguistiques y sont toujours évaluées. S'agissant des connaissances linguistiques, nous estimons que toute évaluation de compétences doit les exclure étant donné que les savoirs linguistiques des apprenants ne sont pas suffisants pour que des énoncés cohérents soient élaborés dans les diverses situations de communication. Ainsi, les questions évaluant les savoirs déclaratifs devraient être écartées pour ne retenir que les questions évaluant réellement la compréhension (QCM, Vrai/faux...) et les situations d'intégration évaluant la production. Pour corriger ces activités, les enseignants devraient recourir à des barèmes (évaluant la compréhension) et à des grilles (évaluant la production). Ces outils devraient aboutir à des

remarques qui précisent le degré de maîtrise notamment que la note serait un outil de jugement peu efficace.

Toutes ces hypothèses que nous tenterons de vérifier tout au long de notre étude nous conduisent à supposer que, pour une évaluation sommative conforme aux prescriptions de l'APC, les enseignants pourraient se contenter d'ajuster leurs pratiques évaluatives actuelles. Mais, à quelles méthodes faut-il recourir pour les vérifier ?

La méthode que nous retenons est l'analyse qualitative et quantitative des épreuves sommatives du cycle moyen qui sont réparties en deux corpus : les épreuves sommatives (devoirs et compositions) et les BEM. Le premier corpus -qui englobe des devoirs et des compositions récents- est analysé en appliquant une grille conçue en nous référant à la littérature scientifique consultée pour une description satisfaisante des pratiques évaluatives actuelles. En vue d'appuyer les résultats auxquels nous parvenons, nous recourons à une analyse des BEM : une analyse comparative confrontant les BEM pré-APC aux BEM post-APC. Cette analyse investit la même grille (outil de collecte) qui est constituée de plusieurs indicateurs extraits des informations exposées dans notre travail sur la validité. Les résultats obtenus sont analysés puis discutés pour que des conclusions soient élaborées.

Suite à cette analyse des corpus collectés, nous proposons des dispositifs évaluatifs (des outils de mesure et de jugement) plus adaptés à la nouvelle approche intégrée par notre système éducatif. Pour ce faire, nous nous référons à des outils existant dans les ouvrages, les guides, les documents pédagogiques...tels les tests conçus par le CECRL que nous estimons plus adaptés que nos épreuves sommatives à une logique de compétences, les grilles proposées dans les manuels...Ces outils que nous proposons seront évalués et validés par des inspecteurs de l'éducation nationale moyennant une liste de vérification inspirée des travaux de Gérard (2009), Roegiers (2004, 2005)...

Cette conception d'épreuves valides constitue le premier objectif de notre étude : étant donné que la nouvelle approche vise à développer des compétences chez les apprenants, les évaluer à travers des pratiques adéquates devient une nécessité. En adaptant ces pratiques évaluatives, les savoirs et savoir-faire -dont la maîtrise est souvent vérifiée à travers des items séparés et décontextualisés- ne seront plus évalués car ils ne sont pas suffisants pour agir dans les situations authentiques de communication. A travers notre premier objectif, nous atteignons un deuxième aussi important : actualiser les pratiques de certains enseignants qui demeurent prisonniers de leurs démarches traditionnelles en évaluation, en dépit de la réforme entreprise. S'agissant de ceux qui ignorent les nouvelles pratiques, il est important de les informer à travers



toutes les précisions que nous puissions dans des recherches et des ouvrages récents de la nécessité d'intégrer ces pratiques pour une évaluation réelle des compétences de leurs apprenants. A ces deux premiers objectifs, s'ajoute un troisième que nous devons mentionner. Il s'agit de la précision des exigences de l'évaluation des compétences : les caractéristiques des outils de mesure, les critères de leur sélection, les caractéristiques des outils de jugement...

Cet intérêt ostensible que nous accordons à la validité des épreuves sommatives au cycle moyen est né de l'écart significatif constaté -durant les années de notre exercice au cycle moyen- entre les résultats obtenus par les apprenants et leurs performances en FLE dans les situations scolaires. Outre cet écart, notons les difficultés qu'ont les enseignants de ce cycle à concevoir des épreuves évaluant les compétences de leurs apprenants ; d'où leur recours fréquent à des examens préfabriqués qui accordent une importance particulière à la vérification de la maîtrise des savoirs grammaticaux enseignés. Ces facteurs nous poussent à étudier les épreuves évaluatives pour répondre aux questions précédemment soulevées.

Nous ne prétendons pas, néanmoins, être les premiers ni les seuls à traiter un sujet si complexe. Un ouvrage récent paru en 2019 qui s'intitule « *Le baccalauréat en Algérie et l'évaluation des compétences- Cas du français langue étrangère* » aborde la validité des épreuves de FLE au BAC ; bien que la validité y soit indirectement traitée à travers l'évaluation de ces épreuves pour déterminer à quel degré elles sont conformes aux exigences institutionnelles. Contrairement à cet ouvrage, la validité est clairement étudiée dans notre travail portant sur un corpus différent : les épreuves sommatives du cycle moyen et les BEM.

Pour étudier la validité de ces épreuves, nous répartissons notre étude en deux volets : un volet théorique constitué de trois chapitres et un volet pratique comprenant le même nombre de chapitres.

Le volet théorique de notre étude sera réparti en trois chapitres. Le premier, qui portera essentiellement sur l'évaluation sommative, répondra à de nombreuses interrogations. Entre autres, citons : les définitions de ce type évaluatif, les frontières qui le séparent des autres formes notamment la forme certificative qui lui est souvent associée et les critères professionnels qui le régissent (la fidélité, la validité, l'objectivité, l'équité...). Ces critères, suffisamment connus par les praticiens, nous les expliquerons de manière détaillée afin de repérer les liens qui pourraient les rapprocher. Si nous leur portons un intérêt particulier dans notre étude, c'est pour mettre en valeur le critère sur lequel nous centrerons notre recherche : la validité.

Comme ils ne sont pas particuliers à l'évaluation des langues étrangères, les éléments énumérés ci-dessus seront traités de manière globale : les définitions que nous aborderons et les références sur lesquelles nous nous appuierons aborderont la question de l'évaluation sommative dans l'univers scolaire, sans la précision d'une discipline particulière. Nous extrairons néanmoins quelques définitions de références portant particulièrement sur les langues étrangères.

Vu que la forme sommative est la plus contestée et critiquée de tous les types évaluatifs, nous préciserons quels regards les divers acteurs impliqués (apprenants, enseignants...) lui portent ; une précision qui s'impose si nous tenons à un changement de vision envers cette forme -qui demeure indispensable dans le milieu scolaire- et à une amélioration des outils dont elle se sert. Mentionnons que les précisions exposées seront précédées de quelques informations que nous jugeons utiles à notre étude. Celles-ci répondront brièvement à des questions liées aux rapports qui unissent évaluation et apprentissage ; aux définitions de l'évaluation, ses fonctions et ses formes ; aux évaluateurs dans le milieu scolaire..., des réponses qui permettront de mieux cerner notre thème.

Bien que les critères professionnels de l'évaluation soient définis dans notre premier chapitre théorique, ils seront traités plus profondément dans notre deuxième chapitre. Certes, la validité -étant la qualité la plus importante (Angoff, 1988) et la plus difficile à assurer en évaluation (Mothe, 1975)- capte essentiellement notre attention : nous la définirons puis nous expliciterons ses liens avec les autres critères retenus : la fidélité, l'objectivité, l'équité... Ces liens sont en fait importants à préciser pour déterminer la conception de la validité que nous adopterons dans notre étude notamment que certains auteurs l'opposent aux autres critères alors que d'autres la relèvent au rang d'une qualité suprême regroupant les autres critères cités.

Toutes ces précisions contribueront à mieux définir la notion de validité et à écarter toute confusion terminologique ; une confusion qui se concrétise dans la diversité des conceptions définitives adoptées par les auteurs. Ainsi, les premiers (Tagliante, 2005) continuent à percevoir la validité comme une capacité des outils évaluatifs à mesurer réellement ce qui est évalué, tandis que les seconds (Kane, 2006) l'associent plutôt à la qualité de l'interprétation des résultats obtenus. Ces deux tendances seront exposées de manière détaillée dans notre deuxième chapitre pour parvenir à une conception précise de la validité puis déterminer ultérieurement les critères et les indicateurs assurant cette qualité dans les épreuves évaluatives.

Nous ne pouvons développer cette notion de *validité* sans aborder celle de *validation*, une étape incontournable dans la conception des outils évaluatifs. Bien que la conception des outils d'évaluation ne soit pas traitée à ce niveau de notre étude, nous tenterons de définir la validation pour éclaircir ce qui la distingue de la validité. Nous exposerons par la suite les différentes procédures qui assurent la validation des épreuves conçues par les enseignants pour évaluer la maîtrise des objectifs visés par les programmes d'études. A travers cet aperçu des procédures de validation, nous déterminerons celles qui seraient appropriées pour la validation des dispositifs intégrés dans le système éducatif algérien pour évaluer les compétences des apprenants en FLE.

Le dernier chapitre théorique sera centré sur les étapes de la conception de dispositifs évaluatifs valides qui nous orienteront lors de la proposition de dispositifs évaluatifs adéquats au contexte national. Mentionnons que nous tenons -à ce niveau- compte des exigences de l'approche adoptée par l'école algérienne (APC). Avant de nous étaler sur les étapes de conception, nous expliquerons ce que nous entendons par *dispositif* en nous référant à la thèse de Roy (2017) intitulée « *Les dispositifs d'évaluation certificative mis en œuvre par les enseignants pour évaluer les attitudes dans un programme d'études collégiales : étude de cas* ». Son travail nous servira également à retracer les phases que les enseignants évaluateurs devront suivre en vue de concevoir des outils évaluatifs mesurant les compétences de leurs apprenants. Les informations que ce chapitre exposera fourniront ainsi des réponses à de nombreuses questions relatives aux objets à évaluer dans une perspective de compétences ; les outils évaluatifs préconisés... Il convient de préciser que ces outils évaluatifs englobent les outils de mesure (exercices, questions, situations...) et de correction (barèmes, grilles...).

Par ce dernier chapitre, nous clôturons notre volet théorique pour entamer la partie pratique de notre recherche, qui sera également constituée de trois chapitres. Le premier sera consacré au cadre méthodologique : nous y expliciterons les méthodes adoptées, les outils de collecte de données investis, le contexte évoqué, le corpus à analyser... Ainsi, pour les méthodes, vu leur multitude, nous sélectionnerons celles qui répondront le plus à nos objectifs, notamment que notre étude ne s'appuiera pas sur une démarche unique. Quant à nos instruments de collecte (une grille d'analyse et une liste de vérification), nous en détaillerons les critères et les indicateurs pour en expliquer l'utilité. Outre ces précisions, nous étudierons le contexte dans lequel s'inscrit notre recherche : nous citerons les évaluations sommatives introduites au cycle moyen ainsi que les moments de leur intégration.

Le chapitre portant sur le cadre méthodologique sera suivi par un chapitre où les résultats collectés seront exposés, analysés puis discutés pour préciser à quel point les épreuves sommatives analysées -constituant notre corpus- sont valides, c'est-à-dire conformes aux exigences de l'APC. Précisons que :

- Les résultats relatifs aux indicateurs seront présentés sans tenir compte de leur ordre dans la grille pour une analyse suffisamment pertinente ;
- Les résultats relatifs aux différents niveaux du cycle seront présentés séparément en cas de différences significatives et indissociablement dans le cas inverse...

Quelle que soit la manière de les présenter, les résultats obtenus seront analysés en parallèle, ce qui permettra de vérifier si chacun des indicateurs exigés pour la validité des épreuves analysées est respecté. Cette analyse sera suivie d'une discussion qui nous permettra de nous pencher sur des études qui approuveraient ou rejetteraient les conclusions primaires formulées. C'est à ce niveau de notre recherche que nous exposerons les résultats de la confrontation des deux catégories de BEM, constituant notre second corpus : les BEM pré-APC et les BEM post-APC. Les conclusions que nous formulerons suite à cette comparaison appuieront les conclusions primaires obtenues suite à l'analyse de notre premier corpus (devoirs et compositions trimestriels des années scolaires : 2017/2018, 2018/2019 et 2019/2020). Grâce aux méthodes investies, nous parviendrons à des conclusions suffisamment fondées et pertinentes qui permettront de comprendre la réalité de l'évaluation scolaire du FLE dans notre contexte ; une compréhension indispensable pour apporter les solutions escomptées aux problèmes éventuels.

Notre dernier chapitre pratique consistera, en réalité, en une contribution qui présente des outils de mesure et de jugement conçus en tenant compte des étapes énumérées par Roy (2017), exposées dans notre volet théorique. Pour les outils de mesure, nous proposerons des situations évaluatives de FLE ; puis nous les validerons en exploitant notre liste de vérification qui comporte des critères et des indicateurs de validité inspirés des références consultées. Quant aux outils de jugement, seule une grille critériée sera proposée pour corriger les productions des apprenants. Cette grille englobera des critères de correction pertinents, indépendants, pondérés et peu nombreux comme le recommandent des auteurs comme Gérard (2005, 2009). Ce sont des critères dont la sélection dépendra de la particularité de la discipline évaluée et des instructions que contiennent les documents pédagogiques. Qu'ils soient minimaux ou de perfectionnement, ces critères seront répartis en indicateurs qui fournissent des indications

## Introduction générale

concrètes sur la maîtrise du critère ; et suivis d'une échelle d'appréciation qui déterminera le moyen à adopter pour exprimer le jugement.

Le dispositif exposé dans notre dernier chapitre ne sera qu'une proposition que des recherches futures pourraient réviser, ajuster et enrichir pour mieux répondre aux besoins du cycle moyen et de notre système éducatif.

# **PARTIE THÉORIQUE**

# **CHAPITRE 1 : AUTOUR DE L'ÉVALUATION SOMMATIVE**

## Introduction

Actuellement, l'évaluation est un thème qui suscite l'intérêt des chercheurs soucieux de développer des dispositifs d'évaluation répondant aux besoins des acteurs concernés. Dans le monde scolaire, cet intérêt est aussi vif que dans le monde professionnel vu les liens étroits qu'entretiennent les deux processus en question : l'évaluation et l'apprentissage. Deux processus indéniablement soudés, ceux-ci ne l'étaient plus auparavant vu le peu d'intérêt que les recherches en éducation lui consacraient. C'est pour comprendre ces liens que nous aborderons dans ce premier chapitre l'évolution du terme évaluation à travers les temps.

En effet, il paraît que la sanction associée souvent à l'évaluation est une conception jugée archaïque, d'où l'abondance des définitions qui tentent de cerner le concept en adoptant des visions plus constructives. Dans ce contexte, nous recensons une multitude de définitions qui n'attribuent plus à l'évaluation des fonctions de sanction ; mais plutôt de diagnostic, de régulation et de remédiation. Ces conceptions sont à l'origine de formes plus utiles en évaluation ; car dans l'univers scolaire, nous assistons à l'introduction de plusieurs types d'évaluation soutenant l'apprentissage dans tous ses moments. Ces formes que nous tenterons de définir à travers notre premier chapitre sont extrêmement utiles pour un meilleur rendement scolaire. Cependant, nous ne pouvons leur consacrer un volet considérable car nous avons choisi de centrer notre étude sur un type particulier : l'évaluation sommative.

A première vue, le choix de ce type d'évaluation semble hâtif dans la mesure où l'évaluation formative est la forme privilégiée récemment dans les interventions enseignantes pour des apprentissages effectifs et durables. Cependant, en considérant les enjeux de l'évaluation sommative et ses retombées qui affectent tous les acteurs pédagogiques et sociaux, ce choix hâtif devient judicieux et fondé. En effet, ses fonctions multiples ne se réduisent pas à la prise de décisions, mais auront des répercussions sur l'orientation des apprenants ainsi que sur leur insertion dans le monde du travail. Contrairement aux autres formes, les résultats des apprenants en évaluation sommative ne déterminent pas seulement les actions entreprises par les enseignants (remédiation) et leurs décisions (le passage à un autre niveau ou le redoublement). Ils sont également décisifs dans l'insertion des évalués dans le marché du travail étant donné que ces résultats servent de descriptifs de leurs compétences dans les différentes disciplines. En considérant certaines situations de difficultés scolaires et de problèmes d'orientation, il semble que les résultats de cette évaluation ne fournissent pas les renseignements nécessaires et ne donnent pas un aperçu réel sur les compétences des apprenants évalués, d'où la nécessité de lui consacrer plus d'attention, et surtout de recherches.



Avant de nous étaler sur la forme sommative de l'évaluation, nous consacrons le premier élément de ce chapitre théorique à l'évaluation dans l'objectif de dessiner les contours de notre étude. En effet, nous tenterons de répondre, même brièvement, à certaines questions qui nous semblent utiles : quels rapports l'évaluation et l'apprentissage entretiennent-ils ? Quelle définition pouvons-nous retenir pour comprendre l'évaluation ? Quelles en sont les fonctions ? Quelles en sont les formes ? Qui sont les évaluateurs dans le contexte scolaire ? Et qu'évaluent-ils ? Nous ne prétendons pas à travers ces questions, auxquelles nous tenterons de répondre, cerner le thème choisi qui demeure d'une complexité indéniable.

Suite aux réponses à cette série de questions, nous passons aux réponses sur des interrogations liées à la forme sommative. Ces interrogations se rapportent à la définition du type, la précision de ses fonctions, la détermination des différences qui le séparent des autres formes notamment la forme certificative qui lui est souvent associée... En plus de ces éléments, nous nous étalons sur les critères professionnels de cette évaluation. Bien qu'ils soient très connus par les praticiens, nous les expliquons de manière détaillée en vue de repérer les liens qui pourraient les rapprocher. Cet élément est certes privilégié dans notre étude car il nous permettra de mettre en valeur le critère sur lequel nous centrerons notre recherche : la validité.

Tous ces éléments seront traités de manière globale vu qu'ils ne sont pas particuliers à l'évaluation des langues étrangères. En effet, les définitions que nous aborderons et les références sur lesquelles nous nous appuyons abordent la question de l'évaluation et de l'évaluation sommative dans l'univers scolaire, sans la précision d'une discipline particulière. Néanmoins, quelques définitions sont extraites de références qui traitent particulièrement de l'évaluation des langues étrangères.

Ayant constaté que la forme sommative est la forme la plus contestée et critiquée parmi les formes répandues (diagnostique, formative et sommative), nous intégrons un élément qui précise quels regards les divers acteurs impliqués lui portent. Qu'ils soient légitimes ou infondés, ces regards ou points de vue doivent être nécessairement compris pour un changement de vision envers la forme sommative. En évoquant ce point-ci, nous ne voulons pas prendre la défense de cette forme contestée, ni lui réclamer une vision valorisante. Nous voulons seulement montrer que l'évaluation sommative -en dépit de toutes les critiques qui lui sont à tort ou à raison attribuées- demeure une forme indispensable car elle est profondément ancrée dans tous les systèmes éducatifs. Au lieu de la contester pour la supprimer, il serait plus fructueux d'adopter une vision plus utile : la critiquer pour améliorer les outils dont elle se sert.

## **1. A propos de l'évaluation**

## 1.1. Deux paradigmes en évaluation

Avant de définir la notion d'évaluation, nous évoquons le changement de sa signification, qui a largement contribué à la construction de sa définition actuelle. L'auteur Ralph W. Tyler (1986) - cité par Howe et Ménard (1993)- évoque dans ce sens un changement de paradigme<sup>1</sup>. En effet, il distingue le paradigme ancien du paradigme nouveau que nous expliquons dans ce qui suit.

### 1.1.1. Le paradigme ancien (le paradigme du test theory)

Les caractéristiques de cet ancien paradigme, où l'évaluation était conçue comme un outil de classement entièrement dissocié de l'apprentissage scolaire, sont énumérées par Guba et Lincoln (1991), cités par Howe et Ménard (1993). Nous les reprenons dans les points suivants :

- Mesure et évaluation : deux notions confondues :

Le paradigme classique se caractérise particulièrement par la confusion de deux notions considérées équivalentes : mesure et évaluation, avec une certaine priorité accordée à la mesure. Les deux notions étaient en effet interchangeable dans la mesure où l'évaluation n'est possible qu'à travers la mesure. Insistant sur le lien entre les deux notions, Thorndike et Hagen (1961) estiment que « *de bonnes techniques de mesure fournissent une solide base pour une évaluation saine* » (Cités dans Nadeau, 1988, p. 44). En d'autres termes, la qualité de l'évaluation dépend de celle de la mesure. Précisions que les deux notions - équivalentes au terme *docimologie*<sup>2</sup> dans la francophonie- « *tenaient davantage de la psychométrie que du programme scolaire et du curriculum* » (Howe & Ménard, 1993, p. 55), ce qui expliciterait une certaine rupture entre apprentissage et évaluation caractérisant cette époque.

- Lien archaïque entre apprentissage et évaluation :

Le deuxième élément qui caractérise le paradigme classique est lié aux rapports qu'entretiennent apprentissage et évaluation. Ainsi, l'évaluation était entièrement dissociée de l'apprentissage pour plusieurs raisons telles que la rupture -mentionnée ci-dessous- entre la

---

<sup>1</sup> Le concept de paradigme est défini par Howe et Ménard, qui se réfèrent à Ouellet (1981), comme suit : « *un ensemble d'hypothèses sur lesquelles on organise et développe des théories et des modèles. C'est un schéma global qui sert de repère dans nos investigations, dans nos recherches* » (1993, p. 54).

<sup>2</sup> Le concept de *docimologie*, associé par certaines études à la mesure et à l'évaluation, renvoie selon G. De Landsheere à une « *science qui a pour objet l'étude systématique des examens, en particulier des systèmes de notation, et du comportement des examinateurs et des examinés* » (2000, p. 17). Cette science est distincte de la mesure étant donné qu'elle assure la validité, la pertinence (De Vecchi, 2011) et la fidélité des outils de la mesure (les examens) et de jugement (les notes ou les échelles de notation).

mesure et le curriculum, et entre la mesure et les objectifs de la mesure. En effet, l'évaluation avait pour seul objectif l'attribution de scores aux évalués pour les classer. Il s'agit dans cette perspective d'une *évaluation normative*<sup>3</sup> qui ne s'appuie pas sur des critères précis. Notons l'absence de la dimension informative qui ne permet plus aux enseignants-évaluateurs le développement de compétences effectives chez leurs apprenants.

- Classement : un objectif dépassé

La deuxième caractéristique du paradigme classique nous conduit à aborder la troisième qui se rapporte à l'objectif visé par les éducateurs et les enseignants à travers leurs tests. Ceux-ci ont constitué, selon Howe et Ménard (1993), un outil pour classer les individus en deux catégories : les individus éducatibles et ceux qui ne le sont pas. Elitiste, la mesure a également assuré la sélection des élèves qui ont les aptitudes requises à la poursuite de leurs études : c'était le rôle attribué aux tests d'intelligence. Outre la répartition des apprenants en catégories, l'évaluation a constitué un outil de recrutement. En effet, les militaires ont été recrutés sur la base des tests proposés, pendant les deux guerres mondiales. Ainsi conçue, l'évaluation - dont la principale fonction est le classement des élèves- est réduite à un outil de sanction qui entrave le progrès des élèves et le développement de leurs compétences.

Vu ses caractéristiques, cet ancien paradigme présentait plusieurs limites d'où son délaissement au profit d'un paradigme plus rénové. D'ailleurs, nous comptons de nombreux auteurs comme Tyler qui ont dénoncé ces limites en s'appuyant sur des arguments judicieux. Dans cette perspective classique de la mesure, une image négative d'une école individualiste – privilégiant la compétition et étouffant la collaboration entre apprenants - est prônée. Ces facteurs contribuent à installer des conditions psychologiques défavorables au développement des compétences chez les apprenants. En effet, la menace que constitue la sanction crée une atmosphère d'insécurité génératrice d'échec. De cette manière, l'évaluation ne permettrait plus à l'école d'atteindre les objectifs réels des différents acteurs pédagogiques (apprenants, enseignants et parents) : le développement de compétences transférables dans des situations authentiques, d'où la montée du nouveau paradigme, plus adapté aux attentes des acteurs pédagogiques.

---

<sup>3</sup> Evaluation normative : selon cette approche, nous classons les apprenants les uns par rapport aux autres. Les résultats sont souvent représentés sous forme de courbes comme la courbe de Gauss qui est souvent critiquée dans le domaine pédagogique, par des auteurs comme Paristot (1990). Dans ce type, la réussite et l'échec ne sont pas conditionnés par la maîtrise des apprentissages ou des compétences acquises par les apprenants : le taux de réussite est stable d'une année à une autre de façon à obtenir le même nombre d'apprenants.

### 1.1.2. Le paradigme rénové (Évaluation liée à l'apprentissage)

Cette rénovation du paradigme en évaluation est due à plusieurs facteurs. Selon Howe et Ménard (1993), le nouveau contexte social, économique et technologique requiert de l'école la formation d'individus compétents et autonomes. Par opposition au paradigme classique, le nouveau présente l'évaluation comme une partie intégrante du processus de l'apprentissage et distincte de la mesure, selon l'affirmation de D'amour (1996) citée par Leroux et Bigras (2003).

Les éléments qui caractérisent le nouveau paradigme sont présentés dans les points suivants :

- Mesure et évaluation : deux notions distinctes

Confondues dans l'ancien paradigme et associées à la docimologie, la mesure et l'évaluation prennent des acceptions complètement différentes. Ainsi, Howe et Ménard se réfèrent à Tyler pour affirmer « *que les concepts de mesure et d'évaluation sont des concepts distincts, des processus différents et que la mesure n'est qu'une façon de réaliser l'évaluation* » (Howe & Ménard, 1993, p. 57). En présentant la mesure comme un procédé évaluatif, l'affirmation de Tyler suscite des interrogations quant aux autres méthodes qui peuvent s'opposer à la mesure en évaluation des compétences. De son côté, Cuq (2003) incorpore la mesure dans les diverses étapes de l'évaluation qui englobe : l'intention, la mesure, le jugement et la décision.

La distinction établie entre mesure et évaluation est évoquée par des auteurs autres que Tyler. Ainsi, Taba (1962), citée par De Landsheere (2000), oppose les deux processus. Elle écrit :

Le processus de mesure est fondamentalement descriptif, car il indique quantitativement à quel point un trait est possédé. La mesure en éducation est généralement concentrée sur des caractéristiques spécifiques, étroites et bien définies. L'évaluation dépend de la mesure, mais elle porte sur un profil plus large de caractéristiques et de compétences<sup>4</sup> (Citée dans De Landsheere, 2000, p. 21).

A travers cette citation, nous repérons deux aspects (descriptif et quantitatif) de la mesure, ce qui en occulte le caractère qualitatif. Cependant, en nous référant à Mottier Lopez, qui avance que la préoccupation de la mesure en évaluation était d'« *assurer une qualité à la notation scolaire* » (2013, p. 941), ce caractère qualitatif devient également intrinsèque à la mesure. Pour ce qui est de l'évaluation, elle se caractérise par son aspect général étant donné

---

<sup>4</sup> Précisons que la citation telle que nous l'avons reprise est attribuée par De Landsheere à Taba H. sans indiquer la page ni la citation originale en anglais.

qu'elle se sert des données de la mesure pour établir un jugement global sur la compétence évaluée.

La distinction établie par H. Taba entre mesure et évaluation, qualifiée d'*utile* par De Landsheere, est acceptable avec une réserve. En effet, De Landsheere réfute le passage de toute évaluation par la quantification qui ne peut pas être valable dans certaines situations telles que les productions complexes (De Landsheere, 2000). Ce point nous amène vers une autre distinction entre les deux concepts que nous retrouvons dans des études plus récentes. Ainsi, Mottier Lopez (2013) a précisé que l'évaluation implique des opérations autres que la mesure telles que l'interprétation des résultats. D'ailleurs, la notion de mesure implique une certaine scientificité quasi-impossible en évaluation. Si la précision de la mesure peut être atteinte en évaluation de certains traits clairement observables, elle ne pourrait l'être en évaluation de traits<sup>5</sup> plus complexes. De ce point de vue, la notion de mesure semble compatible avec certaines disciplines aux contenus aisément évaluables (les mathématiques, les sciences...). C'est pour cette raison que la notion est notamment investie dans des études réalisées par des scientifiques. Dans ce contexte, Reuchlin (1962), auteur auquel Porlier et Langouet (1991) se réfèrent, rappelle la définition du physicien anglais Campbell (1938), qui conçoit la mesure comme une estimation chiffrée de la valeur des objets évalués et insiste ainsi sur le caractère quantitatif de la notion. Ce caractère quantitatif et chiffré de la mesure est également mis en valeur par des chercheurs comme Guilford et Fruchter (1973), cités par De Landsheere (2000), dont la définition rejoint celle de Campbell. L'appréciation numérique ou la notation, caractéristique de la mesure, n'est pas toujours requise en évaluation scolaire. Dans ce cas-ci, nous tenons à préciser qu'elle est réfutée -voire écartée- par certains auteurs (Hadji, 2015) car elle est perçue comme un outil de dévaluation.

- Évaluation : une partie intégrante de l'évaluation

Dans le nouveau paradigme, évaluation et apprentissage entretiennent de nouveaux rapports : l'évaluation sert l'apprentissage, selon l'expression de Meirieu (1989), de Hadji (1990) et de Gérard (2013). Après une rupture avec le paradigme classique, l'évaluation devient étroitement associée à l'apprentissage. Elle ne se réduit plus à une appréciation du degré de la maîtrise des compétences ciblées chez les évalués pour les classer.

---

<sup>5</sup>Le concept de trait consiste en la capacité d'agir face aux diverses situations. Il se rapproche de celui de compétence.

Grâce à ce nouveau paradigme, les fonctions de l'évaluation sont bien déterminées. En effet, celle-ci intervient au début de l'apprentissage pour la planification de contenus conformes aux objectifs fixés (évaluation diagnostique), pendant l'action pédagogique pour la régulation des dysfonctionnements constatés et au terme des apprentissages pour la prise de décisions. Ainsi intégrée, l'évaluation devient une partie intégrante de l'apprentissage. Dans ce sens, Hadji (1990) évoque un apprentissage assisté par l'évaluation.

En somme, les affirmations de Meirieu et de Hadji abordées ci-haut insistent sur l'importance de l'évaluation, un processus qui apporte aux enseignants une aide considérable pour une intervention éducative efficiente.

- Autonomie : un nouvel objectif

Le nouveau paradigme, où l'évaluation est en lien étroit avec l'apprentissage, rompt avec les objectifs du paradigme classique. Dans celui-ci, l'évaluation a souvent constitué un outil de classement des apprenants selon leurs résultats (scores) obtenus : elle était normative. Cependant, dans le paradigme rénové, l'école à travers une évaluation plutôt *critériée*<sup>6</sup> vise la formation d'apprenants autonomes, ce qui est dû particulièrement à la prolifération des sources de savoirs. En effet, les apprenants ont accès, grâce aux nouvelles technologies d'information et de communication, à une multitude de sources d'informations. Pour les assimiler, ils sont contraints à agir sans le guidage de leur enseignant. Ce sont ces situations auxquelles tout enseignant doit préparer ses apprenants. Pour atteindre cet objectif (l'autonomie des apprenants), l'évaluation doit être investie à bon escient. Elle doit vérifier la maîtrise des objectifs ciblés par les programmes retenus.

Nous précisons à ce niveau de notre étude que ces changements des objectifs et des rapports, que l'évaluation et l'apprentissage entretiennent, s'expliquent par les rénovations pédagogiques dont la plus considérable est la création des curriculums. Cette création, nous la devons à Tyler qui a « *suggéré d'organiser les curriculums autour d'objectifs de formation* » (Howe & Ménard, 1993, p. 57), ce qui conduit à la conception d'instruments évaluatifs permettant de vérifier si les objectifs institutionnels, définis par les curriculums, sont atteints (Howe & Ménard, 1993). Nous devons reconnaître que cette création des curriculums

---

<sup>6</sup> L'évaluation critériée est la forme où la compétence manifestée par les apprenants, ou performance, est évaluée en fonction de critères déterminant les exigences attendues (seuil de réussite) pour ladite compétence. Selon Ouellet, cette forme « *contrairement à l'évaluation normative, est personnalisée et est en référence à des critères, des objectifs* » (1983, p. 375). En effet, chaque étudiant est évalué selon les mêmes critères que ses camarades, mais ses résultats dépendent des critères évaluatifs réussis, qui correspondent au niveau minimum pour atteindre l'objectif visé.

a fondé les bases d'une évaluation valide dans la mesure où les items évaluatifs correspondent aux objectifs institutionnels.

En réalité, la rénovation du paradigme en évaluation, qui avait plusieurs mérites tels que la détermination de nouvelles fonctions évaluatives (la régulation...), est certes dû au changement du contexte social, économique, culturel et technologique. Cependant, nous ne pouvons pas nier le rôle que les enseignants-évaluateurs ont joué dans la mesure où leur attitude déterminerait la réussite de toute innovation pédagogique.

## **1.2. Vers une définition de l'évaluation**

Ce qui vient d'être abordé (les deux paradigmes en évaluation) constitue une esquisse à la définition du concept-clé de notre étude : évaluation. Cependant, son apposition à la mesure nous paraît insuffisante pour en cerner toutes les acceptions. En effet, une définition satisfaisante de ce processus complexe doit tenir compte des moments sur lesquels celui-ci s'étale ainsi que des instruments à mobiliser. Dans ce qui suit, nous citons les éléments sur lesquels les définitions insistent dans leurs définitions pour aboutir à une définition plus globale.

### **1.2.1. Évaluation : un jugement de valeur**

Dans sa conception classique, l'évaluation était liée à la mesure qui consiste en la quantification des performances évaluées. Cependant, celle-ci est réfutée car une évaluation est plutôt qualitative : une appréciation ou un jugement. Selon Scallon, Barbier (1983) considère le jugement de valeur comme une particularité de l'évaluation. En effet, ce processus est souvent associé à une attribution de valeur à ce qui est évalué. Dans ce contexte, le Conseil supérieur de l'éducation du Québec avance que « *l'évaluation en éducation consiste à déterminer la valeur de quelque chose* » (1983, p. 12, cité dans Scallon, 1988, p. 18). Cette deuxième définition, à première vue, écartée du domaine éducatif en raison de la non-précision des objets évalués, met en exergue une première acception de l'évaluation associée à un jugement de valeur.

Cette perspective est également adoptée en didactique de FLE. Ainsi, J-P. Robert définit le concept dans son « *Dictionnaire pratique de didactique du FLE* ». Celui-ci « *consiste à mesurer, à juger des performances les compétences orale et écrite des apprenants placés dans des situations simulées de communication* » (2008, p. 84). Certainement la plus utile à notre étude qui s'inscrit dans le domaine de la didactique des langues étrangères, la définition de J-P. Robert, qui englobe la mesure, a plusieurs avantages. D'une part, elle précise les objets évalués par les enseignants-évaluateurs en langue : la performance. Il est connu que la compétence n'est pas observable et que son évaluation porte sur sa manifestation : la

performance. D'autre part, elle fournit des informations sur les situations d'évaluation que doit contenir un test de langue. Outre ces deux avantages, cette définition est indéniablement la mieux ancrée dans la nouvelle approche actuellement appliquée dans plusieurs systèmes éducatifs : l'Approche Par les Compétences, dorénavant l'APC.

Cette première conception qui associe l'évaluation à un jugement de valeur ou appréciation ne pourrait être aisément validée : elle prête à de nombreuses objections. En premier lieu, elle véhicule une connotation subjective de l'évaluation étant donné que tout jugement ne peut être exempt de subjectivité. Certes, la subjectivité, souvent critiquée, est inévitable en évaluation comme l'affirme Durand qui écrit : « *le jugement est un acte professionnel qui ne peut revêtir un caractère de totale objectivité* » (2008, p. 66, cité dans Ndinga, 2011, p. 124). En second lieu, le jugement professionnel<sup>7</sup> ne peut être exempt de biais (effet de halo, effet de contamination, la fatigue du correcteur...) qui en affectent la validité (Ndinga, 2011).

De cette manière, définir l'évaluation en tant que jugement (qui provient d'une seule personne) présente plusieurs limites auxquelles le recours à plusieurs juges semble constituer une solution, bien que la concertation de ceux-ci ne puisse être assurée dans tout contexte. Toutefois, le jugement évoqué s'appuie sur des informations recueillies par différents moyens, d'où les définitions qui insistent sur la phase de la collecte des informations.

### **1.2.2. Évaluation : collecte des informations**

Par opposition à certaines définitions qui associent l'évaluation à un jugement de valeur, celles que nous proposons dans la deuxième catégorie insistent notamment sur la collecte des informations. Ce sont ces informations recueillies qui permettent de prendre les décisions appropriées fondées sur une interprétation valide.

Dans son ouvrage « *L'évaluation formative des apprentissages* », Scallon définit l'évaluation comme « *un processus par lequel on délimite, obtient et fournit des informations utiles permettant de juger des décisions possibles* » (Stufflebeam & coll. 1980, p. 48, cités dans Scallon, 1988, p. 19). Le recueil des informations, dont il est question dans cette citation, vise à prendre des décisions appropriées. Il nous semble important de préciser que ces informations ou données sont recueillies (par des observations, appréciations ou résultats) lors de la mesure (Cuq, 2003). Des auteurs comme Bloom et ses collaborateurs préfèrent plutôt évoquer la

---

<sup>7</sup> Le jugement professionnel est un processus qui s'appuie sur la collecte de données en vue de prendre des décisions judicieuses, respectant les principes de l'équité et de déontologie (Allal & Lafortune, 2008).



collecte de preuves. Ainsi, l'évaluation consiste en « *la collecte systématique de preuves pour déterminer si certains changements ont effectivement lieu chez les apprenants et pour déterminer la quantité ou le degré de changement chez chaque élève* » (Scallon, 1988, p. 19)<sup>8</sup>. La tendance définitoire adoptée par ces auteurs est largement inspirée de la conception argumentative en évaluation. Ainsi, toute évaluation est un processus qui consiste à recueillir des preuves sur les compétences des évalués. Ces compétences, comme la citation ci-dessus l'avance, se manifestent dans des changements qui seront quantifiés et qualifiés chez chaque apprenant.

Cependant, une évaluation ne se conclut pas par la collecte des données. En effet, celles-ci seront interprétées pour aboutir à des jugements justifiés, d'où la nécessité d'intégrer les la collecte des informations et le jugement, des moments interdépendants, dans les définitions de l'évaluation.

Si la collecte des informations (preuves), dont il est question dans les définitions précédentes, concerne les informations relatives à la compétence des évalués, des auteurs tels que Noel-Jothy et Sampsonis (2006) insistent sur la dimension informative de toute évaluation qui doit fournir des informations réelles et utiles aux apprenants, aux enseignants et aux parents. De ce point de vue, ces auteurs insistent plutôt sur une fonction d'évaluation, celle d'informer les acteurs pédagogiques.

Outre ces définitions qui insistent sur la phase de la collecte des données en évaluation, certaines définitions, que nous aborderons dans le point suivant, mettent en avant les finalités de l'évaluation : les décisions.

### **1.2.3. Évaluation : notation et prise de décision**

Toute évaluation aboutit également à des décisions qui se concrétisent dans des notes, des annotations... Dans de nombreux systèmes éducatifs, l'opération évaluative se conclut par l'attribution d'une note qui reflète le degré de maîtrise de la compétence évaluée. Dans cette optique, Médioni affirme qu'« *Evaluer dans une conception « ordinaire » largement répandue à l'école signifie en effet noter et classer* » (2016, p. 11). Cette citation tout en établissant un lien entre évaluation et notation en montre le caractère ordinaire et restreint étant donné que l'évaluation est un processus plus large qui ne pourrait être réduite à la notation (Hadji, 2010).

---

<sup>8</sup> Traduction de « *the systematic collection of evidence to determine whether in fact certain changes are taking place in the learners as well as to determine the amount or degree of change in individual students* » (Bloom & coll. 1981: 5, cités dans Scallon, 1988, p. 19).

En réalité, la notation n'est qu'un outil pour représenter un écart entre ce qui doit être maîtrisé et ce qui est réellement maîtrisé par tout apprenant évalué. En dépit de toutes les critiques souvent assignées à la notation comparée à une *torture pédagogique*<sup>9</sup>, cette méthode considérée comme un outil pour informer les divers acteurs (Tagliante, 2005) subsiste en dépit de toutes les rénovations pédagogiques.

Des décisions découlent de cette conception de l'évaluation en tant que notation. En fonction des notes attribuées aux évalués, les décisions de réussite et d'échec sont prises par les acteurs ayant la prérogative de les émettre (enseignants ou institution). Outre les décisions de réussite et d'échec, la notation continue à conditionner les orientations des étudiants à différents niveaux, dans quelques systèmes éducatifs. Ce qui expliquerait les lacunes qu'éprouvent beaucoup d'étudiants qui se retrouvent démunis dans des filières qui ne correspondent pas à leurs compétences effectives.

Les résultats des apprenants en évaluation ne déterminent pas seulement leur réussite ou leur échec, leur orientation... Ils fournissent des informations utiles aux enseignants pour la planification de programmes de régulation visant la remédiation aux lacunes constatées. Cette fonction régulatrice est souvent citée dans les définitions que nous reprenons dans ce qui suit.

#### **1.2.4. Évaluation : un processus de régulation**

Cette conception que nous avons ajoutée est justifiée par de nombreuses définitions selon lesquelles l'évaluation est une partie intégrante de l'apprentissage. Celle-ci « *s'inscrit de plus en plus en dynamique avec une logique régulatrice, formative, formatrice (pas) uniquement de vérification contrôle* » (Ardouin, 2013, p. 215). De ce point de vue, nous pouvons intégrer Hadji qui considère que « *l'évaluation est intelligente dans la mesure où elle participe à son niveau à la progression des élèves dans la construction de leurs connaissances et compétences* » (Hadji, 2015, p. 12). Cette tendance met l'accent sur le rôle régulateur de l'évaluation et sur sa fonction de remédiation assurée notamment par un type particulier : l'évaluation formative. En dépit de ses avantages indéniables, l'évaluation formative ne doit pas écraser l'existence d'autres types d'évaluation tels que l'évaluation sommative dont le poids est plus important en réalité.

En somme, toutes les définitions que nous venons de regrouper sous quatre catégories insistent sur un des aspects intégrés en évaluation. Cependant, ces catégories ne sont pas

---

<sup>9</sup> Nous empruntons cette expression à Peter Gumbel, cité par Hadji, dans son ouvrage : *Évaluation à l'école : pour la réussite de tous les élèves* (2015).

exhaustives dans la mesure où des auteurs comme Tyler (1942) et De Ketele (1989) ont pris en considération d'autres aspects tels que l'écart entre performance et objectif. Ainsi, Tyler (1942) conçoit l'évaluation comme une « opération qui mesure l'écart entre un objectif prédéterminé que l'on poursuit et le résultat obtenu » (Bonniol, 1886<sup>10</sup>, cité dans Ardouin, 2007, p. 76)<sup>11</sup>. Pour sa part, De Ketele (1989) évoque l'écart entre un « ensemble d'informations et un ensemble de critères adéquats aux objectifs fixés au départ ou ajustés en cours de route, en vue de prendre une décision » (Cité dans Gérard & Roegiers, 2009, p. 120). Ces deux auteurs sont unanimes sur la conception de l'évaluation en tant que comparaison de la performance manifestée par l'apprenant évalué à la performance attendue, exprimée sous forme de critères. Contrairement aux définitions qui ne privilégient qu'un des aspects du processus évaluatif, la définition proposée par Cuq (2003) englobe toutes les phases du processus (intention, mesure, jugement, décision), ce qui la rend la plus complète. Nous tenons à préciser dans cet ordre d'idées que l'élaboration d'une définition qui ne soit pas contestée n'est pas une tâche aisée dans la mesure où elle doit tenir compte de l'intention de l'évaluateur, de l'objet évalué et du moment de l'évaluation.

De tout ce qui précède, nous pouvons affirmer que l'évaluation est un processus large qui englobe diverses phases : la préparation (conception des épreuves), la passation (épreuves soumises aux évalués pour recueillir les données), la notation et l'interprétation (correction des copies et attribution de notes ou de remarques), le jugement (précision du degré de maîtrise des compétences par les évalués) et la décision (passage ou redoublement). Ainsi, nous rejoignons Roegiers qui affirme que « l'évaluation est (...) un processus qui commence quand on se fixe un objectif à atteindre, et qui se termine quand on a pris une décision en rapport à cet objectif » (Gérard & Roegiers, 2009, p. 128). Pourtant générale, cette définition renvoie à divers types d'évaluation qui nous semblent indispensables à expliquer, dans ce qui suit.

### 1.3. Types d'évaluation

Évoquer les types de l'évaluation nous semble nécessaire pour cerner leurs caractéristiques et développer le type particulier (évaluation sommative), qui constitue le thème de notre étude. De nombreux ouvrages énumèrent ces types. Parmi les ouvrages consultés, « *Un Cadre Européen Commun de Référence Pour les Langues : Apprendre, enseigner, évaluer* »

---

<sup>10</sup> La page du travail de Bonniol n'est pas citée par Ardouin.

<sup>11</sup> L'auteur Ardouin, que nous avons cité dans notre recherche, ne cite pas Tyler (1942).

constitue à nos yeux la référence la plus pertinente car nous y trouvons une liste non exhaustive opposant divers types. Nous reprenons ci-dessous le tableau cité dans cet ouvrage :

1	Évaluation du savoir	Évaluation de la capacité
2	Évaluation normative	Évaluation critériée
3	Maîtrise	Continuum ou suivi
4	Évaluation continue	Évaluation ponctuelle
5	Évaluation formative	Évaluation sommative
6	Évaluation directe	Évaluation indirecte
4	Évaluation de la performance	Évaluation des connaissances
8	Évaluation subjective	Évaluation objective
9	Évaluation sur une échelle	Évaluation sur une liste de contrôle
10	Jugement fondé sur l'impression	Jugement guidé
11	Évaluation holistique ou globale	Évaluation analytique
12	Évaluation par série	Évaluation par catégorie
13	Évaluation mutuelle	Auto-évaluation

**Tableau 1: Types d'évaluation**

**(Un cadre européen commun de référence : apprendre, enseigner, évaluer, 2001, p. 139).**

A ces types que le CECRL liste, nous ajoutons d'autres types comme l'évaluation diagnostique (prédictive), l'évaluation formatrice, l'évaluation certificative, l'évaluation ipsative, la co-évaluation, l'évaluation formaliste, l'évaluation naturaliste ... Tous ces types s'opposent selon des paramètres particuliers comme le moment d'évaluation<sup>12</sup>, les objets évalués<sup>13</sup>, les outils évaluatifs<sup>14</sup>, les évaluateurs... Précisons que ce dernier paramètre permet une distinction entre : l'auto-évaluation, l'évaluation mutuelle, la co-évaluation et l'hétéro-évaluation. Deux précisions devraient être avancées sur cette typologie-ci. La première est que ces derniers types listés sont imbriqués par Scallon (1988), Morissette (2010) et Vial (2001) dans les modalités de l'évaluation formative. La deuxième est qu'ils ne sont plus considérés comme des types, mais plutôt comme des formes dans la mesure où certains auteurs tels que Tripet-Lièvre (2017) ont distingué dans leurs études *types* et *formes*, ce qui n'est pas le cas dans notre étude où nous parlerons indifféremment de types et de formes.

<sup>12</sup> Ce paramètre permet une distinction entre les évaluations diagnostique, formative et sommative.

<sup>13</sup> Ce paramètre oppose l'évaluation du savoir, de la capacité, de la performance et de la connaissance.

<sup>14</sup> Grâce à ce paramètre, nous distinguons l'évaluation directe de l'évaluation indirecte.

De tous les types cités, les évaluations diagnostique, formative et sommative sont les plus connues et intégrées en éducation. Cette intégration dans de nombreux systèmes éducatifs est due à la particularité de ces types qui renvoient à des fonctions évaluatives tout en recouvrant tous les moments de l'apprentissage scolaire.

### **1.3.1. Évaluation diagnostique (prédictive)**

Le premier type que nous évoquons est l'évaluation diagnostique ou prédictive qui intervient avant les apprentissages. Plus récente selon Scallon (1988), cette forme vise à déterminer le profil d'entrée des apprenants. Par profil d'entrée, nous désignons tous les savoirs, les savoir-faire et les compétences qu'ils maîtrisent au début de l'année scolaire. Outre ce qu'ils maîtrisent, cette évaluation détermine également leurs difficultés. En effet, à travers ce qu'ils ne maîtrisent pas, des décisions relatives aux modalités, aux outils et aux stratégies à mobiliser seront prises. Dans ce contexte, Gérard précise que ce type d'évaluation « *permet de déterminer les forces et les faiblesses de l'élève et de vérifier si l'élève maîtrise les savoirs, savoir-faire et compétences nécessaires et préalables à l'apprentissage. Cela permettra ainsi d'orienter l'enseignement de façon plus adaptée* » (2009, p. 49). Cette citation qui affirme nos propos montre la visée informative du diagnostic pédagogique qui recense les connaissances et les compétences des apprenants pour un objectif double : vérifier si ce qui est maîtrisé correspond à ce qui est requis pour entamer les nouveaux cours d'une part, et adapter l'enseignement prévu aux besoins réels. A première vue, l'appellation évaluation prédictive paraît distante de celle d'évaluation diagnostique. Cependant, elle semble justifiée si nous prenons en considération la double fonction qui lui est assignée par le CECRL : donner un aperçu sur les compétences de chaque élève et prédire le niveau qui serait atteint par l'enseignement prévu.

En dépit de son importance, nous mentionnons le manque de références consacrées à cette forme évaluative en éducation, ce qui peut être justifiée par son apparition récente. Cependant, nous considérons que cette forme est plutôt intégrée en formation dans la mesure où le programme d'études dans celle-ci est élaboré en tenant compte des besoins recensés, ce qui n'est pas le cas dans l'enseignement scolaire où le programme retenu ne tient pas compte de l'hétérogénéité des apprenants ni de leurs besoins effectifs. Ainsi, des études devaient-elles porter sur cette forme évaluative particulière afin d'en montrer l'application dans les trois cycles du système éducatif algérien qui la préconise.

Dans le domaine de la formation, l'évaluation diagnostique semble plus intégrée. Selon Hadji (2000), cette forme vise l'orientation des formés vers des filières, qui correspondent à

leurs compétences réelles, et l'adaptation de la formation et des contenus sélectionnés à leurs profils.

### 1.3.2. Évaluation formative (régulatrice)

Considérée comme la plus importante, cette forme d'évaluation est intégrée au moment de l'apprentissage pour le réguler et remédier à toutes les lacunes recensées. Les auteurs tels que Hadji (2010) qui ont attribué à l'évaluation une fonction régulatrice accordent, même indirectement, à l'évaluation formative un certain privilège.

Tout comme le premier type évoqué, celui-ci consiste en un recueil d'informations déterminant ce qui n'est pas maîtrisé au cours de l'enseignement dispensé, autrement dit les erreurs commises. Ainsi conçue, cette forme évaluative accorde à l'erreur un nouveau statut, celui que privilégient les approches constructivistes. Dans ce contexte, Scallon (1988) affirme que dans cette forme introduite par Scriven en 1967, théoriciens et pédagogues sont invités à réfléchir en adoptant des positions différentes sur l'erreur qui constitue un outil de remédiation. Dans cet ordre d'idées, l'auteur explicite la démarche formative qui consiste à vérifier par des tests et des activités si les objectifs visés sont atteints. Cette vérification est suivie d'activités pour la remédiation aux lacunes constatées.

En dépit des avantages souvent associés à cette évaluation formative, celle-ci n'est pas exempte de désavantages. En effet, elle « *présente l'inconvénient de mettre l'élève en situation d'évaluation constante (Bosman)* » (Paquay ; Defèche & Dufays, 2002, p. 93), créant ainsi une atmosphère d'insécurité défavorable à l'apprentissage scolaire. De cette manière, l'évaluation qui constitue une partie intégrante et une aide à l'apprentissage devient un obstacle à celui-ci. De ce fait, il devient primordial de rétablir la fonction principale de l'évaluation formative, celle d'informer et d'intégrer tous les acteurs du processus enseignement/apprentissage pour assurer un développement plus réussi des compétences.

Nous tenons à préciser que certains auteurs tels que De Landsheere (1979) et Scallon (1988) associent ce deuxième type au troisième que nous évoquerons ci-dessous (évaluation sommative). Ainsi, De Landsheere établit-il une distinction entre les deux types. Pour lui, « *l'évaluation formative revêt un caractère privé (alors que) l'évaluation sommative est publique* » (De Landsheere, 1979, p. 115). Par opposition à l'évaluation sommative, l'évaluation formative, plus centrée sur l'apprenant, n'intègre pas d'autres acteurs tels que les parents. Quant à Scallon, il estime que l'évaluation formative à travers la remédiation aux lacunes constatées chez les apprenants les prépare à l'évaluation sommative. Celle-ci sera développée dans le point ci-dessous.

### 1.3.3. Évaluation sommative

Ce troisième type que nous évoquerons est selon Médioni « *la modalité de l'évaluation la plus connue et usitée* » (2016, p. 13). Intégrée au bout d'une séquence, d'un trimestre ou d'un semestre d'apprentissage, l'évaluation sommative vérifie le degré d'assimilation des contenus enseignés, contrôle la maîtrise des compétences développées et classe les évalués (Hadjji, 2015 ; Scallon, 1988).

L'évaluation sommative est associée à des concepts tels que contrôle des apprentissages (Scallon, 1988) et bilan (Lussier, 1992). Elle se réalise à travers des activités évaluatives variées (exercices, tâches, situations...) dont la complexité est conditionnée par plusieurs variables (niveau des apprenants, volume des apprentissages, approche adoptée...). En dépit de la nécessité de tenir compte de certaines variables dont ceux cités ci-dessus, certaines activités évaluatives persistent dans l'univers scolaire comme les exercices structuraux auxquels les enseignants recourent souvent en évaluation (Huver & Springer, 2011). Le maintien de ces activités dévoile une certaine réticence des enseignants-évaluateurs qui demeurent fidèles à des outils évaluatifs dont la validité est contestée en raison de leur incapacité à révéler les compétences réelles des apprenants. Ces activités aboutissent à une note qui devrait refléter les apprentissages maîtrisés par les apprenants. Cette note est selon Tagliante (2006) l'unique information donnée. Certes, la note est un outil réfuté car elle ne fournit aucuns renseignements sur la compétence évaluée. Cependant, elle continue à être un outil privilégié, voire indispensable, dans beaucoup de systèmes éducatifs. Nous tenons à préciser qu'une accumulation d'évaluations sommatives, selon l'expression de Mègre et Riba (2015), est possible au cours de l'année scolaire. Elle se concrétise dans des interrogations, des tests, des examens périodiques, des contrôles, des épreuves...

Par opposition aux autres formes évaluatives, l'évaluation sommative est souvent confondue avec un autre type : évaluation certificative. Ainsi, en assignant à l'évaluation sommative les fonctions de vérifier les apprentissages et d'attester la maîtrise des compétences ciblées, des auteurs tels que Hadji (2010) prennent les deux évaluations sommative et certificative pour une seule. Si certains auteurs prennent les deux formes pour une seule, d'autres « *affirment qu'il est souvent utile de distinguer l'évaluation certificative de l'évaluation sommative afin de réserver la certification à l'essentiel* » (Paquay ; Defêche & Dufays, 2002, p. 92). Cette distinction nous semble indispensable dans la mesure où une évaluation certificative aboutit à l'attribution d'une certification (un diplôme), contrairement aux évaluations sommatives.

Imposée institutionnellement, l'évaluation sommative est certainement la forme la plus socialisée étant donné que les résultats des apprenants sont communiqués à tous les acteurs sociaux. A ce caractère socialisé et informatif, s'ajoutent ses retombées dans l'orientation et la réussite ou l'échec scolaire.

Étant donné que l'évaluation sommative consiste à vérifier si les objectifs visés sont atteints, il est nécessaire de déterminer les objectifs sur lesquels celle-ci porte. En effet, il est difficile d'évaluer tous les contenus dans le cadre d'un examen dont la durée est limitée. Dans cet ordre d'idées, il est important de mentionner que dans le cadre de l'APC, des situations évaluatives permettant une évaluation valide des compétences visées sont proposées. Celles-ci doivent être conformes aux situations d'apprentissage, avec de nouveaux contextes (Huver & Springer, 2011).

En somme, nous pouvons affirmer que les trois formes d'évaluation, détaillées dans notre étude, se distinguent par leurs moments, leurs fonctions et leurs outils. Ainsi, l'évaluation diagnostique est intégrée en amont des apprentissages pour une adaptation des contenus retenus aux particularités des apprenants. Pour l'évaluation formative, elle est intrinsèque à toutes les phases de l'apprentissage afin de remédier aux erreurs constatées. Enfin, l'évaluation sommative aboutissant à une note est intégrée en aval des apprentissages pour vérifier la maîtrise des compétences visées. C'est à cette dernière forme particulière que la suite de notre étude sera consacrée.

#### **1.4. Les évaluateurs dans le contexte éducatif**

Qui évalue dans le contexte éducatif ? Une telle question semble de prime abord insignifiante dans le cadre de notre recherche centrée sur la validité de l'évaluation sommative. Cependant, elle devient significative pour identifier certains types d'évaluations (Huver & Springer, 2011) qui se rapportent à la forme sommative. A ce premier avantage, s'ajoute la précision, apportée à travers une telle question, des rôles impartis aux acteurs du processus évaluatif.

##### **1.4.1. Évaluateurs distincts des formateurs**

Dans un premier cas, les concepteurs des épreuves ou les évaluateurs pourraient être des personnes distinctes des enseignants. A ce niveau, il est question de l'évaluation externe où les évaluateurs sont « *des personnes qui n'appartiennent pas au dispositif de formation* » (Huver & Springer, 2011, p. 97). Ce cas-ci correspond dans notre système éducatif algérien aux évaluations sommatives pratiquées à la fin de chaque cycle (primaire, moyen, lycée) où les épreuves proposées aux apprenants sont conçues par des personnes distinctes de celles qui



assurent leur enseignement. Une multitude de tests sont mis à la disposition des organismes évaluateurs en vue de sélectionner des apprenants, recruter des employés... Ces tests sont souvent couronnés par un diplôme qui « *jouissant d'une reconnaissance nationale ou internationale permet à son détenteur de se positionner sur un échelon de la société de façon incontestée* » (Noel-Jothy & Sampsonis, 2006, p. 8). En effet, toutes les institutions sont mandatrices de certifications validant les compétences réelles des publics concernés.

#### **1.4.2. Enseignants-évaluateurs**

Dans un deuxième cas, l'évaluation est assurée par les enseignants qui sont chargés de l'enseignement. Ainsi, ces enseignants conçoivent les épreuves évaluatives, élaborent les barèmes de correction et corrigent les copies des apprenants. De ce point de vue, l'évaluation est considérée comme une tâche de l'enseignant. Selon Gérard, « *l'évaluation des apprentissages fait partie de l'enseignement et à ce titre, les enseignants se doivent de mettre en place un dispositif non seulement pertinent, valide et fiable, mais aussi transparent et loyal* » (2009, p. 193).

Il est connu que tout enseignant a des tâches à assurer en classe. Celles-ci ne se réduisent plus à la transmission de savoirs et de savoir-faire chez les apprenants. En effet, l'enseignant doit sélectionner les outils évaluatifs valides, fiables et pertinents pour une évaluation objective dont les fonctions sont multiples. Ainsi, il évalue le profil d'entrée de ses apprenants afin de déterminer leurs besoins authentiques, ce qui assure une adaptation des contenus retenus. Tout au long de son intervention, il doit veiller à une vérification du degré de maîtrise de ce qui est enseigné, ce qui permet de prévoir des activités de remédiation appropriées. Enfin, il évalue les compétences de ses apprenants au terme de son intervention pédagogique afin de donner un aperçu fiable, pertinent et valide sur ce qu'ils maîtrisent.

Dans le cadre de notre travail, ce sont les évaluations sommatives conçues par les enseignants qui constituent une partie de notre corpus dont l'étude nous permettra de recenser les pratiques les plus courantes pour en étudier la validité.

#### **1.4.3. Apprenants-évaluateurs**

Dans ce dernier cas, qui correspond à l'auto-évaluation, l'apprenant-évaluateur doit « *vérifier les étapes de sa démarche d'apprentissage et (...) stimuler ses progrès sur des points précis* » (Peretti & al. 1998, p. 534, Cités dans Huver & Springer, 2011, p. 97). L'auto-évaluation est considérée comme une compétence cruciale par de nombreux auteurs tels que Bélair et Coen (2015). Cette habileté requiert cependant un certain niveau dans la mesure où

elle est une dimension de l'autonomie St-Pierre (2004). Bélair (2015) s'est intéressée à l'évaluateur-évalué selon divers points de vue. Du point de vue pédagogique, l'auteure évoque la complexité de ce processus étant donné que « *pour pouvoir juger de sa performance, de sa pratique ou encore de la qualité de ses productions, la personne doit développer un esprit critique de niveau supérieur* » (Bélair, 2015, p. 40). Ce niveau-ci exige un certain recul pour éviter toute imitation du formateur, d'où le recours à des outils tels que les grilles, les questionnaires... La citation nous apprend que l'auto-évaluation ne porte pas uniquement sur les idées ou les travaux comme nous le croyons souvent. Elle doit tenir compte des démarches adoptées par les apprenants. En effet, ceux-ci doivent analyser leur itinéraire lors de la résolution des tâches ou au cours des intégrations de nouveaux savoirs. Cette analyse leur permet d'extraire des informations sur leurs styles, stratégies et rythmes d'apprentissage, éléments qui orientent leurs futurs apprentissages. Cependant, nous rappelons encore que l'auto-évaluation requiert certaines compétences car elle est une dimension de l'autonomie et une habileté métacognitive (St-Pierre, 2004). Rejoignant St-Pierre dans sa conception de l'auto-évaluation, Scallon (1988) insiste sur la nécessité de développer cette compétence chez les apprenants dans la mesure où ils auront à l'investir dans leurs apprentissages extrascolaires. St-Pierre, qui se réfère à Scallon (2004) avance que « *le développement de la capacité à s'autoévaluer est une tendance importante en éducation. L'auto-évaluation est, selon lui, une habileté à développer, en d'autres termes un savoir-faire à s'approprier* » (2004, p. 34). Dans cet ordre d'idées, St-Pierre (2004) a abordé, dans son article intitulé « L'habileté d'auto-évaluation : pourquoi et comment la développer ? », les pratiques enseignantes qui pourraient amener au développement de cette compétence réflexive.

A travers la présentation des acteurs qui assurent l'opération évaluative, il s'avère que le rôle de l'évaluateur n'est plus exclusif à l'enseignant chargé de la formation. Il pourrait être assuré par l'apprenant lui-même, qui adopte un regard critique sur son propre progrès afin de réguler sa démarche, son activité et sa production, ou par un évaluateur externe pour répondre une demande exprimée par les institutions concernées (entreprises, universités...). Chacun des acteurs supra cités constitue un pilier dans la réussite du processus d'évaluation. Cependant, nous estimons que l'enseignant est l'acteur auquel une attention particulière devrait être accordée. En effet, celui-ci assure la réussite de la formation des évalués, mais les amène progressivement à devenir autonomes et à s'autoévaluer. Ces tâches à remplir par l'enseignant exigent une formation sérieuse en matière de l'évaluation scolaire.

### 1.5. Fonctions de l'évaluation

Selon plusieurs auteurs (Perrenoud, 2001 ; Cuq, 2003), les fonctions de l'évaluation correspondent aux diverses décisions prises par les enseignants en classe. Elles varient selon les chercheurs. Ainsi, Blanchet (2010), dans « *Réformer l'évaluation scolaire : mission impossible ?* », aborde les fonctions connues et souvent évoquées (formative, sommative ou certificative, pronostique) auxquelles Vögeli-Mantovani (2000) ajoute l'évaluation de l'institution, tandis que Weiss (1998)<sup>15</sup> propose la fonction informative. Pour sa part, De Landsheere (1976) recense les rôles suivants de l'évaluation : l'inventaire, le diagnostic et le pronostic. Dans le même sens, Bélair (1999) associe les fonctions proposées aux types distingués en fonction du moment de l'évaluation (évaluation diagnostique, formative et sommative). Ainsi, au premier type, les fonctions orienter et adapter sont attribuées. L'évaluation formative, deuxième type, a pour fonctions de réguler et de faciliter. Enfin, le dernier type a les rôles suivants : vérifier et certifier. Perrenoud (2001) distingue trois fonctions, étroitement liées aux décisions à prendre par l'enseignant. Il s'agit de la fonction formative, certificative et pronostique. Ces trois fonctions correspondent à celles que Blanchet a citées supra. Cardinet (1983) attribue à l'évaluation les trois fonctions suivantes : la régulation, la certification et l'orientation. Strittmatter (2007) estime que la classification la plus originale et utile des fonctions de l'évaluation est celle que Nisbet (1990) a proposée. En raison du manque des ouvrages qui évoquent cette classification, nous nous limitons à citer les typologies établies par les auteurs cités supra.

Nous constatons que les fonctions formative et sommative sont présentes dans toutes les typologies proposées par les auteurs, ce qui en démontre l'importance dans le milieu scolaire. Ce sont ces deux fonctions que Cuq (2003) retient dans son « *Dictionnaire de didactique du français langue étrangère et seconde* ». D'autres fonctions, que Bélair (1999) a citées, semblent aussi importantes que les deux précédentes. En effet, la fonction informative et l'orientation sont des fonctions indispensables dans toute évaluation institutionnelle.

### 1.6. Qu'évalue-t-on dans le milieu scolaire ?

La dernière question à laquelle nous répondons dans cet élément de notre chapitre est : qu'évalue-t-on dans le milieu scolaire ? Cette question nous permettra de déterminer les acteurs

---

<sup>15</sup> Ces deux auteurs sont cités dans l'ouvrage *Réformer l'évaluation scolaire : mission impossible ?*, édité par Patricia Gilliéron Giroud et Ladislas Ntamakiliro (2010).

à évaluer dans notre étude notamment qu'en éducation, il est possible d'évaluer les apprenants, les établissements, les dispositifs, les enseignants... (Huver & Springer, 2011).

Dans notre étude, l'évaluation porte sur les apprenants : nous évaluons leurs compétences. Une distinction devrait être établie entre l'évaluation des apprenants et celle de leurs compétences. Contrairement à la première qui est inutile et subjective, la deuxième est réclamée par tous les acteurs pédagogiques. L'évaluation des élèves est ainsi un biais qui affecte la validité et la fiabilité de ses résultats : une connaissance antérieure de ceux-ci ou de leurs compétences pourrait subjectiver la correction de l'enseignant. Par évaluation des apprenants, nous entendons l'évaluation des performances qui se manifestent dans des projets réalisés, des énoncés produits, ses représentations sémantiques...

Le deuxième acteur de la classe sur lequel porte l'évaluation est l'enseignant. Selon Paquay, « *lorsqu'on évalue un enseignant, on recueille des informations à propos de certaines caractéristiques de cet enseignant et on les compare avec les caractéristiques attendues* » (2004, p.15). Ces caractéristiques ne sont pas précises dans la mesure où elles dépendent du point de vue de l'évaluateur (apprenant, parent...). Cependant, des critères sont retenus comme la qualité de l'enseignement (Paquay, 2004) et les résultats des apprenants (Smith, 2003 ; Laveault, 2007). Soulignons que les enseignants sont évalués à travers des pratiques formelles (inspecteurs, directeurs...) ou informelles (apprenants, parents...). Précisons que les résultats des apprenants servent aussi à évaluer la classe (une classe est jugée bonne si les résultats de ses apprenants le sont), voire l'institution (les résultats des apprenants inscrits dans un établissement donné en fondent la réputation).

L'évaluation dans le contexte scolaire porte également sur les contenus qui sont souvent révisés pour une meilleure adaptation aux besoins des apprenants et aux exigences de leur vie authentique, notamment que la qualité des contenus détermine les compétences des apprenants. Elle porte aussi sur les démarches et les stratégies déployées en classe vu son rôle à « *vérifier que les méthodes didactiques et la pédagogie ont une action efficace sur l'apprenant* » (Veltcheff & Hilton, 2003, p. 08, Cités dans Huver, Springer, 2011, p. 98).

Pour conclure cet élément, nous précisons que nous limiterons notre étude à l'évaluation sommative des compétences des apprenants en langue étrangère (le français dans le système éducatif algérien).

## **2. Évaluation sommative : définitions**

De toutes les formes, cette évaluation est la plus connue et intégrée (Médioni, 2016). Elle est également la plus ancienne : la fonction sommative est intrinsèque au concept évaluation dès son existence, ce qui n'est pas le cas des fonctions formative (introduite par Scriven en 1967) et diagnostique (introduite par Bloom, Hastings et Madaus en 1971). Contrairement à ces deux types qui lui sont souvent opposés, l'évaluation sommative est une pratique obligatoire dans de nombreux systèmes éducatifs (Prégent, 1990), d'où l'importance que nous lui accordons. A travers cet élément, nous tenterons de la définir en nous appuyant sur ses caractéristiques listées par Prégent (1990) et par Bélec (2017), qui synthétisent les travaux d'Aylwin (1995), Leroux (2001) et Scallon (2000). En effet, selon Prégent, cette forme se distingue par le moment de son intégration, son but, sa rétroaction et son cadre de référence. Certes, nous ne nous limiterons pas à ces caractéristiques et nous inclurons toutes celles qui s'y rapportent.

### **2.1. Évaluer au terme des apprentissages**

Le moment de l'intégration de cette évaluation est un critère fondamental qui assure sa distinction des autres formes (diagnostique et formative). En effet, l'évaluation sommative intervient au terme des apprentissages scolaires, constituant ainsi le bilan des acquis des apprenants.

Plusieurs auteurs tels que Prégent (2000), Cuq (2003), Pelpel (2005), Ardouin (2013)... insistent sur cet élément dans leurs définitions. Ainsi, Cuq estime que l'évaluation sommative est intégrée « *à la fin d'un cours, d'un cycle, d'un programme d'études ou d'une partie terminale de programme* » (2003, p. 91). Elle est introduite alors au terme des apprentissages, qui pourraient être enseignés au cours d'un semestre, d'une séquence, voire d'un cours. Ce dernier cas nous conduit à nous interroger sur la possibilité de son exploitation à des fins régulatrices. Contrairement aux autres moments (fin du programme, du cycle...) qui supposent un état d'avancement considérable et entravent toute rétroaction remédiant aux lacunes constatées, la fin du cours semble un moment propice pour un retour réflexif sur ce qui est appris. En effet, tout comme l'évaluation sommative qui intervient au terme du cours, l'évaluation formative est intégrée, selon Danvers, qui se réfère à De Landsheere (1979), « *au terme de chaque tâche d'apprentissage et ayant pour objet (...) de découvrir où et en quoi un élève éprouve des difficultés d'apprentissage, en vue de lui proposer ou de lui faire découvrir des stratégies qui lui permettent de progresser* » (1994, p. 114). Nous ne nous étalons pas sur ce lien éventuel entre évaluation sommative et formative que nous étudierons infra. Rejoignant Cuq (2013), Tardieu et Ardouin définissent l'évaluation sommative en prenant en considération

le critère du moment de son intégration. Néanmoins, contrairement à la conception de Cuq, Tardieu (2008) estime que ce bilan doit être établi après un ensemble de tâches qui pourraient correspondre à une séquence d'apprentissages ou un trimestre. Il en est de même pour Ardouin (2013) qui situe cette évaluation au terme d'une période considérable d'apprentissages étant donné que le bilan requiert le cumul des apprentissages. Cependant, cette période considérable ne correspond pas forcément à une année scolaire dans la mesure où certains auteurs à l'instar de Mègre et Riba (2015) évoquent la possibilité de multiplier les évaluations sommatives au cours d'une année, d'où les questions qui s'imposent sur le nombre idéal d'évaluations sommatives à intégrer dans le milieu scolaire généralement, et en classe de FLE particulièrement. Soulignons que Tardieu, qui -à l'instar des auteurs cités supra- situe l'évaluation sommative en aval des apprentissages, associe celle-ci à la « *somme de compétences* »<sup>16</sup>, une conception ancrée dans une perspective d'évaluation de compétences.

A ce premier critère (le moment d'évaluation) qui permet de définir la forme sommative, s'ajoutent ses buts qui sont considérablement distincts de ceux des autres formes. Nous les développerons dans ce qui suit.

## 2.2. Buts particuliers à l'évaluation sommative

Toute forme d'évaluation est intégrée à des fins précises. Bien que la fonction régulatrice lui soit assignée (Bélec, 2017), l'évaluation sommative se distingue par son but : le contrôle de la maîtrise des connaissances et des compétences développées chez les apprenants, à travers diverses activités. Les résultats obtenus dans ces activités informent les apprenants et leurs enseignants, et permettent de distinguer ce qui est maîtrisé de ce qui ne l'est pas (Tardieu, 2008). Dans cet ordre d'idées, nous pouvons évoquer le rapport étroit mentionné par des chercheurs tels que Scallon (1988) et Lussier (1992) entre cette forme d'évaluation et des concepts comme contrôle et bilan. En plus de son premier but, cette évaluation vise la prise de décisions relatives aux différents acteurs pédagogiques et sociaux.

Certaines décisions prises par les enseignants concernent en premier lieu la qualité des apprentissages que les apprenants assimilent, selon Prégent (1990), dans la mesure où une évaluation devrait révéler ce qui est réellement maîtrisé par celui-ci de ce qui ne l'est pas. Autrement dit, les compétences réellement maîtrisées par les apprenants sont déterminées de façon à avoir un aperçu transparent sur le profil de chaque apprenant. En quoi ce profil pourrait-

---

<sup>16</sup> Expression employée par Claire Tardieu (2008) dans son ouvrage *La didactique des langues en 4 mots-clés : Communication, Culture, Méthodologie, Evaluation* pour définir l'évaluation sommative, une expression qui révèle une perspective récente en évaluation scolaire et adaptée aux approches en vigueur.

il être utile ? Il nous semble justifié de rappeler à ce niveau de notre travail les enjeux de l'évaluation sommative qui franchissent les contours du milieu scolaire pour atteindre le milieu professionnel où l'accès au marché du travail est sans conteste conditionné par les compétences des candidats, certifiées scolairement, dans un domaine précis. Le profil de tout apprenant a pour ainsi dire des répercussions sur son avenir scolaire et professionnel.

La seconde catégorie de décisions qui relèvent des tâches des enseignants se rapportent au passage à un niveau supérieur. En effet, les résultats obtenus par l'apprenant suite à une évaluation sommative déterminent son passage vers un niveau supérieur ou son redoublement. Dans ce cas, Prigent affirme que « *le professeur joue le rôle d'un censeur qui détermine si la somme des connaissances acquises par un étudiant au terme d'une étape d'un cours est suffisante pour qu'on lui accorde au moins une note de passage* » (1990, s.p<sup>17</sup>). Cette conception de l'évaluation semble étroitement inspirée du paradigme ancien dans lequel l'évaluation, entièrement dissociée de l'apprentissage scolaire, sanctionne les apprenants et les classe selon leur niveau.

Aux deux catégories de décisions relatives au métier de l'enseignant, s'ajoutent les décisions institutionnelles liées aux enseignants et aux établissements. Nous avons précédemment précisé que les résultats des apprenants ne sont pas seulement révélateurs de leurs niveaux et indicateurs de leurs compétences. Ils sont exploités à d'autres fins comme : évaluer les enseignants, évaluer les écoles, évaluer les programmes... Comme la compétence de l'enseignant est déterminée par la réussite de ses apprenants, l'évaluation sommative devient à l'origine de décisions qui le concernent (promotion ou dégradation), qui se rapportent au programme (changement de programmes ou de manuels, réforme de la pédagogie...)... A ce niveau, le contexte national, qui a connu l'intégration de l'Approche Par les Compétences suite aux résultats décevants de la Pédagogie Par Objectifs qui l'a précédée, nous paraît le plus parlant.

Nous tenons à préciser que la prise de décisions associée à l'évaluation sommative amène de nombreux chercheurs à la confondre avec l'évaluation certificative (Hadji, 2015), une confusion que nous n'adopterons pas dans le présent travail où les divergences entre ces deux types en question seront précisées.

---

<sup>17</sup> Les pages de ce livre, que nous avons consultées sur le site Google Books, ne sont pas numérotées.

### 2.3. Un jugement définitif

Le troisième critère cité par Prégent (1990) pour différencier l'évaluation sommative des autres types est la rétroaction. Pour cet auteur, cette forme aboutit à un jugement irrévocable et définitif. En effet, «*le professeur ne souhaite aucunement favoriser un retour en arrière de l'étudiant pour que celui-ci reprenne une partie de son apprentissage en vue de l'améliorer et de bonifier sa note finale* » (Prégent, 1990. s.p.<sup>18</sup>). Contrairement à l'évaluation formative qui suppose un retour rétroactif pour la remédiation et le renforcement des apprentissages, l'évaluation sommative ne correspond à aucune rétroaction régulative. Elle semble ainsi écartée du processus d'apprentissage.

Nous pouvons déceler cette absence de rétroactions en évaluation sommative dans l'affirmation de Jorro qui déclare que «*la mesure conduit à poser des assertions qui ont valeur de jugements, à produire des actes de parole définitifs (Ricœur, 1995)*» (2000, p. 63). Cette citation, comme nous l'avons déjà avancé, rejoint la précédente au niveau du caractère définitif des jugements émis suite aux résultats obtenus. Cependant, elle englobe des éléments qui méritent d'être éclaircis. Le premier élément qui interpelle notre attention est le lien établi entre évaluation et mesure. Ce lien qui remonte au paradigme archaïque de l'évaluation semblerait contestable dans la mesure où actuellement l'évaluation ne se réduit plus à la mesure des connaissances. Néanmoins, nous devons préciser que le paradigme récent, où évaluation et apprentissage sont deux processus étroitement liés et complémentaires, n'exclut pas le paradigme ancien étroitement lié à l'évaluation sommative.

Ce premier élément nous conduit au second que nous développerons dans ce qui suit. Ainsi, la mesure, que nous admettons dans ce contexte en tant que synonyme d'évaluation sommative, s'achève par l'attribution de résultats ou de scores aux travaux évalués des apprenants. Les enseignants traduisent ces notes jugeant la qualité et la quantité des acquis en des annotations ou des jugements évaluatifs. Ce sont ces jugements qui clôturent une séquence, un trimestre ou un semestre d'apprentissage scolaire. En effet, les lacunes constatées et les dysfonctionnements repérés ne seront pas objets d'une remédiation pédagogique pour améliorer les connaissances ou les compétences visées par l'intervention enseignante. Dans ce sens, Prégent affirme que «*le succès ou l'échec des étudiants dans une épreuve sommative donnée change bien peu de choses dans le déroulement ultérieur des cours* » (Prégent, 1990, s.p.<sup>19</sup>).

---

<sup>18</sup> La page consultée sur Google Books n'est pas numérotée.

<sup>19</sup> Page non indiquée.



Ce troisième critère nous semble discutable si nous prenons en considération le nombre d'évaluations sommatives intégrées au cours d'une année scolaire et le type de progression adopté. Si cette évaluation sommative se situe en aval d'un cours, sa fonction devient la régulation (Bélec, 2017) et la rétroaction est inévitable. Il en est de même lorsque cette évaluation est intégrée en aval d'une séquence d'apprentissages. Cependant, intégrée en aval d'un trimestre ou un semestre, un retour réflexif sur les objectifs qui ne sont pas atteints par les apprenants est extrêmement difficile. Cette première variable, qui porte sur le nombre d'évaluations au cours des apprentissages, correspond notamment aux progressions linéaires de contenus. Contrairement à celles-ci, les progressions spirales (appelée également spiralées ou en spirale) sont incontestablement favorables aux rétroactions régulatrices : la répétition des contenus assure leur assimilation progressive par les apprenants. En effet, il est connu que « *progression "en spirale" permet à l'élève de revenir plusieurs fois sur la même notion au cours de la formation, lui laissant ainsi le temps de la maturation, de l'assimilation et de l'appropriation* » (Programme de mathématiques et de sciences physiques et chimiques pour les baccalauréats professionnels, 2009).

Contrairement aux auteurs cités supra qui nient toute rétroaction en évaluation sommative, Bélec (2017) mentionne l'existence des rétroactions dans ce type évaluatif visant l'explication des résultats aux apprenants. Cependant, ces rétroactions n'affectent pas la pratique enseignante dans la mesure où elles n'entraînent pas un retour sur les apprentissages insuffisamment maîtrisés.

#### **2.4. Un cadre de référence double**

Contrairement à l'évaluation formative toujours critériée, l'évaluation sommative se caractérise par un cadre de référence double où deux approches normative et critériée (ou critérielle selon Cuq (2003)- sont distinguées (Scallon, 1988). Selon Prigent, « *l'évaluation sommative n'a pas toujours un caractère normatif. Elle peut avoir lieu dans un cadre de référence critériée* » (1990, s.p.)<sup>20</sup>.

Dans la première approche (critériée), les apprenants sont évalués en fonction de critères précis. Cette approche vise « *à déterminer le niveau de performance langagière atteint par un élève face à des objectifs d'apprentissage visés par des programmes d'études* » (2003, p. 90). Contrairement à l'approche normative, l'approche *critériée* ou *critérielle* est plus

---

<sup>20</sup> Page non indiquée (Google Books).

préconisée pour deux raisons essentielles. La première est que les critères d'évaluation renvoient aux descripteurs de la compétence évaluée, ce qui assure une évaluation valide des acquis des apprenants. Il est connu que la validité des résultats d'une évaluation est conditionnée par celle des instruments investis. Ceux-ci englobent les outils de mesure ou les activités évaluatives (items ou questions), qui doivent être représentatives des objectifs enseignés, et les outils de jugement ou les grilles critériées dont les critères traduisent les attentes des enseignants. En recourant aux critères, nous assurons la validité du jugement et celle de l'évaluation. La seconde raison, plus constructive que la première, concerne la possibilité d'investir les résultats de cette évaluation à des fins formatives. En effet, l'évaluation à référence critériée a l'avantage d'informer les enseignants et les apprenants de ce qui est entièrement maîtrisé, de ce qui est en cours d'apprentissage et de ce qui reste à apprendre. De telles informations conduisent les enseignants à réguler leurs interventions et leurs apprenants à mieux atteindre les objectifs définis. Certes, ce cas particulier d'évaluation sommative se clôt par des jugements (Prégent, 1990).

Pour l'évaluation sommative à référence normative, les apprenants sont classés par rapport à une norme (Cuq, 2003). En effet, selon Prégent (1990), les enseignants recourent à cette évaluation sommative pour le classement des apprenants du plus fort au plus faible, et inversement.

Nombreux sont les auteurs qui se sont intéressés aux avantages de ces deux approches de l'évaluation sommative. Les plus éminents, que nous citons en nous inspirant de Scallon (1988) sont Robert L. Ebel, qui évoque les avantages de l'approche normative, et W. James Popham, qui cite ceux de l'approche critériée. Leurs travaux ont cependant abouti à des résultats inverses dans la mesure où ils ont contribué à la clarification de leurs limites regroupées dans la citation suivante de Scallon :

Il est clair que l'interprétation normative ne convient pas à une épreuve de rendement centrée sur un objectif pédagogique et conçue pour décrire avec précision ce qu'un élève est capable de faire. Il est également évident que l'approche critériée ne convient pas à un examen composé de tâches diverses et conçu pour apprécier le rendement général d'un élève à un registre étendu d'habiletés et de compétences (1988, p. 90).

## **2.5. Notation : exclusive à l'évaluation sommative**

Aux particularités de l'évaluation sommative citées ci-dessus, s'ajoute la notation. Celle-ci reflète les compétences ou connaissances des apprenants évalués. Rejetée en raison de sa valeur informative contestée, la note demeure introduite et interprétée dans les pratiques

pédagogiques. Parmi les auteurs qui attachent la notation à cette forme sommative, nous citons Tardieu qui avance :

L'évaluation sommative (...) fait l'objet d'une notation transmise notamment aux parents. La notation chiffrée doit utiliser l'échelle des notes dans son intégralité : le professeur n'hésitera pas à accorder la note maximale à un exercice ou une activité s'il satisfait entièrement aux critères retenus. Si les notes obtenues sont constamment faibles, il sera nécessaire de revoir le plan et le rythme de progression (2008, p. 208).

Trois éléments dans cette citation captivent notre attention car nous les trouvons révélateurs de la conception évaluative de l'auteur. Le premier est que la notation chiffrée constitue un outil informatif destiné aux parents des apprenants. Cette affirmation est discutable étant donné que tous les acteurs pédagogiques et sociaux réclament ces résultats scolaires pour diverses fins. Selon Tagliante, « *la note a une fonction d'information : elle sert à rendre compte de l'état des élèves, aux parents (bulletin scolaire) et à la société (admission à l'examen)* » (2005, p. 14). Le deuxième élément que la citation éclaire est la possibilité du recours à tous les degrés de l'échelle de notation. Contrairement à certaines pratiques enseignantes qui varient les notes des apprenants entre deux limites définies, Tardieu (2008) n'exclut pas la possibilité d'accorder une note maximale si la performance évaluée répond aux exigences déterminées. Le troisième élément correspond à la nécessité d'une rétroaction régulatrice en cas de résultats insatisfaisants. Ce point de vue rejoint celui des auteurs qui accordent à l'évaluation sommative une fonction de régulation (Bélec, 2017).

En plus de la note attribuée pour juger la qualité de la performance indicatrice de la compétence des évalués, certains enseignants recourent à des annotations ou des appréciations qui fournissent plus de renseignements. Ces écritures évaluatives, qui ne sont pas mieux approuvées que la note, créent cependant quelques difficultés aux enseignants qui s'en servent (Jorro, 2000).

Sans nous étaler sur les critiques assignées à la notation que nous développerons dans la suite de ce travail, nous rappelons le recours à cinq critères fondamentaux pour définir l'évaluation sommative : le moment de son introduction, ses buts, sa rétroaction, son cadre de référence et sa notation. Normative ou critériée, l'évaluation sommative est alors une opération introduite en aval des apprentissages cumulés, pour émettre un jugement sur la compétence atteinte par les apprenants durant une période déterminée (un cours, une séquence, un semestre...). Cette définition qui récapitule les définitions citées, nous la renforcerons par la

distinction, présentée dans ce qui suit, de cette évaluation sommative des autres types certificatif et formatif.

### 3. Croisement des formes : réalité ou mythe ?

En définissant l'évaluation sommative à travers ses caractéristiques, nous avons évoqué la similitude établie par quelques auteurs (Hadji, 2015) entre la forme sommative et certificative, deux concepts dont ils se servent indifféremment. Etant donné que les deux évaluations interviennent au bout de la formation pour aboutir à un jugement, elles sont à raison confondues. En plus de l'équivalence entre ces deux premières formes, il s'est avéré que les résultats de l'évaluation sommative peuvent dans certains cas générer une action régulatrice pour le renforcement des compétences à développer chez les apprenants. Cette fonction régulatrice est essentiellement intrinsèque à l'évaluation formative distincte de la forme sommative, ce qui légitime tout lien entre les deux formes. Tenant compte de ces éléments, nous abordons dans notre étude un certain croisement des trois formes citées vu que certaines de leurs fonctions se ressemblent. Cependant, ce croisement est-il réel ? Est-il justifié ?

#### 3.1. Évaluation sommative : pourrait-elle devenir formative ?

Avant de préciser les liens qui pourraient exister entre les deux formes en question, nous devons préciser que leur opposition n'est pas admise par de nombreux auteurs, tels que De Ketele (2006) et Figari (2006), étant donné que l'évaluation formative renvoie à une finalité alors l'évaluation sommative se rapporte plutôt à une démarche<sup>21</sup>. Selon De Ketele, une telle opposition « *a conduit à restreindre la démarche sommative à la fonction certificative, la démarche descriptive à la fonction régulatrice et formative, la démarche herméneutique ou intuitive<sup>22</sup> à la fonction d'orientation* » (De Ketele, 2006, p. 103, Cité dans Mottier Lopez, Laveault, 2008, p. 18), une restriction qui n'est pas toujours valide dans la mesure où une évaluation certificative pourrait assurer un rôle formatif.

Dans le cadre de notre recherche, cette distinction souvent opérée entre l'évaluation sommative, intégrée en aval des apprentissages pour dresser un bilan des acquis des apprenants, et l'évaluation formative, dont l'intégration au cours de l'apprentissage vise à l'améliorer, est

---

<sup>21</sup> Les points de vue de De Ketele (2006) et de Figaro (2006) sur l'opposition entre évaluation sommative et formative sont abordés par Mottier Lopez et Laveault (2008) dans leur article « L'évaluation des apprentissages en contexte scolaire : Développements, enjeux et controverses ».

<sup>22</sup> Contrairement à la démarche herméneutique globale qui cherche à valider une parmi une série d'hypothèses sur le rendement des apprenants, la démarche descriptive décrit les compétences maîtrisées de celles qui ne le sont pas, offrant ainsi aux enseignants la possibilité d'apporter des régulations appropriées.

maintenue. La fonction de régulation évoquée pourrait cependant exister en évaluation sommative (Bélec, 2017), d'où le croisement des deux formes. Dans ce sens, Noel-Jothy et Sampsonis (2006) avancent que l'évaluation sommative devient parfois formative si les informations collectées, suite au processus évaluatif, visent la rectification de la pratique enseignante pour des apprentissages efficaces.

En abordant le lien éventuel entre les deux formes, nous n'avons pas l'intention de remettre en question la distinction irrévocable entre elles. Il s'agit essentiellement d'insister sur une idée évoquée ci-dessus : la fonction de régulation et de rectification existe même en évaluation sommative. Le croisement des deux formes évaluatives (sommative et formative) est dû ainsi à une fonction partagée : la régulation.

### **3.2. Une distinction indispensable de la forme certificative**

Les rapports entre les deux formes (sommative et certificative) peuvent être qualifiés d'équivoques. Si le CECRL (2011) oppose une évaluation sommative à une autre formative, certains auteurs tels que Paquay, Defêche et Dufays (2002), Mottier Lopez et Laveault (2008) et Gérard (2013) recourent au concept d'évaluation certificative, pour remplacer celui d'évaluation sommative. Cette substitution opérée peut être repérée dans les définitions que ces auteurs ont données à l'évaluation certificative. Ainsi, Gérard, qui en admet l'intégration à la fin des cycles, avance qu'« *il peut y avoir des certifications en cours d'apprentissage : chaque fois qu'un ensemble d'objectifs a été travaillé, l'enseignant peut mettre en place un processus d'évaluation certificative. L'objectif de celle-ci est de prouver que l'élève maîtrise un certain nombre d'acquis* » (Gérard, 2013, p. 77). Cette définition comprend deux certifications : la première est intégrée au terme des cycles pour délivrer des attestations ou des certificats aux évalués, tandis que la seconde est introduite au terme d'une série d'apprentissages pour certifier ce qui est maîtrisé. Cette seconde conception, relative à l'évaluation sommative, conduit Scallon à recourir au concept évaluation certificative (Bloom ; Hastings & Madaus, 1971).

Si les auteurs évoqués ci-dessus préfèrent le concept évaluation certificative à celui d'évaluation sommative, d'autres tels que Jorro (2000), Cuq (2003), Cuq et Gruca (2005), Mègre et Riba (2015) ainsi que Tardieu (2008) continuent à employer le second concept. Ainsi, Cuq (2003), contrairement à De Ketele (2006) qui évoque une démarche sommative et une fonction certificative, oppose une fonction sommative à la fonction formative, reconnue unanimement par tous les auteurs cités. Nous rejoignons cette seconde catégorie de chercheurs qui ne réduisent plus la fonction sommative à une appellation archaïque (Mottier Lopez &

Laveault, 2008) ou identique à l'évaluation certificative (Noel-Jothy & Sampsonis, 2006), un concept employé dans la francophonie pour désigner l'évaluation sommative (Shepard, 2006).

Une distinction entre ces deux évaluations (sommative et certificative) est cependant indispensable dans la mesure où elles ont des objectifs distincts en dépit de la similitude éventuelle de leurs instruments évaluatifs. Ainsi, une évaluation sommative « *fait un bilan des connaissances ou compétences à valeur institutionnelle* » (Hadji, 2015, p. 38) à travers des interrogations, des examens... Certes, ces instruments aboutissent à une note qui reflète les connaissances, les compétences ou les habiletés des apprenants évalués. En évaluation certificative, l'objectif visé est la validation officielle des compétences (Hadji, 2010), d'où l'attribution de diplômes et de certificats attestant ces compétences. De ce fait, l'évaluation certificative devrait être réservée aux examens de fin de cycles qui valident un socle de connaissances et de compétences. Ce qui n'est pas le cas pour l'évaluation sommative qui pourrait vérifier la maîtrise d'une compétence ou un savoir-faire. Cette évaluation est introduite au cours d'un cycle : au terme d'une séance, une séquence, un trimestre, un semestre ou une année.

De tout ce qui précède, nous précisons que, à l'instar de De Ketele (2006), nous admettons le pouvoir régulateur de la fonction sommative d'évaluation. Ainsi, le bilan de connaissances et de compétences établies servira certainement aux enseignants à revoir certaines de leurs pratiques, méthodes et démarches. Cependant, nous insistons sur la nécessité de distinguer la forme sommative de la forme certificative dans la mesure où la certification n'est qu'un cas particulier d'évaluation sommative dans lequel le bilan (notes, scores...) aboutira à une prise de décision (obtention de diplôme, passage à un niveau supérieur...). Dans ce sens, Galiana avance que « *l'évaluation certificative est un type particulier d'évaluation sommative destinée à certifier les acquis en vue de l'attribution d'un diplôme* » (2014, p. 29). Déterminer les liens entre l'évaluation sommative et les deux formes (formative et certificative) délimite les contours de l'évaluation sommative, thème de notre recherche, ce pour la mieux définir.

#### **4. Professionnalisme de l'évaluation sommative : quels critères ?**

Employé par Noel-Jothy et Sampsonis (2006), le concept de *professionnalisme* se rapporte à la définition de critères précis et valides qui régissent la conception des instruments de mesure, la gestion des conditions de passation, la formation des correcteurs, l'information des acteurs concernés... A travers ces critères, « *définis dans les années 1950, lors de la période appelée psychométrique-structuraliste par D. Lussier et Carolyn E. Turner (1995), au moment*

où, (...) le domaine de l'évaluation prenait autant en compte l'outil de mesure que le contenu des épreuves » (Riba & Mègre, 2014, p. 61-62), l'évaluation sommative aboutit à des données réelles et à des informations utiles. En effet, « toute évaluation doit apporter de réelles informations utiles tout à ceux qui apprennent qu'à ceux qui les aident à apprendre » (Hadji, 2015, p. 30) : les informations destinées aux apprenants les aident à mieux atteindre les objectifs visés tandis que celles destinées aux enseignants contribuent à réguler leurs pratiques pour de meilleurs résultats. A ce niveau, nous tenons à préciser, en premier lieu, que ces critères que nous exposerons ci-dessous ne sont pas exclusifs à la forme sommative de l'évaluation (Mègre & Riba, 2015). La seconde précision que nous ajoutons est la différence entre les critères d'évaluation et les critères de correction. Contrairement aux premiers généraux, les seconds, particuliers, sont définis en fonction de l'objet évalué, des objectifs visés, de la matière enseignée et notamment de la situation d'évaluation proposée... Quels sont ces critères professionnels qui assurent l'utilité de toute évaluation ?

Les critères retenus et expliqués dans le reste de notre travail (validité, fiabilité/fidélité, objectivité, faisabilité, équité) sont ceux rassemblés par Noel-Jothy et Sampsonis (2006).

#### **4.1. La validité**

De tous les critères professionnels, la validité est « la qualité à la fois la plus importante et la plus difficile à atteindre pour n'importe quelle épreuve d'évaluation » (Mothe, 1975, p. 23). Cette importance est le motif de ce travail qui vise la vérification de la validité des instruments évaluatifs du système éducatif algérien. La citation de Mothe, qui conçoit la validité comme une qualité du test relative à son objectif (Lado, 1961), expose la perspective adoptée par de nombreux auteurs tels que Buckingham et al. (1921), Kelley (1927), Nunnally (1978), Tagliante (2005), Noël-Jothy et Sampsonis (2006), Riba et Mègre (2015) ... Ceux-ci sont unanimes sur la définition suivante de la validité : la capacité du test à évaluer réellement la compétence évaluée. Comment cette capacité se manifeste-t-elle ?

L'adéquation entre ce que le test vise à mesurer et ce qu'il mesure (Laveault & Grégoire, 2002), ou validité, est déterminée par plusieurs indicateurs dont la correspondance entre les items du test et le contenu enseigné. Cet indicateur est cité par Cardinet qui estime que tout test « doit correspondre aux objectifs du curriculum enseigné » (1990, p. 2, Cité dans de Pierto & Roth, 2017, p. 32). Mothe (1975), de son côté, énumère les indicateurs tels que l'adéquation et la représentativité du contenu et des objectifs du test par rapport aux objectifs de l'apprentissage.

Contrairement à cette perspective classique qui persiste, la perspective récente introduit une nouvelle conception de la validité. Celle-ci n'est plus une qualité du test, mais de l'interprétation des résultats auxquels il aboutit (Cronbach, 1971). Les critiques n'ont pas épargné cette nouvelle conception de la validité dans la mesure où certains auteurs à l'instar de Borsboom, Mellenbergh et Van Heerden (2004) préfèrent à la validité de l'interprétation celle du test. Nous trouvons ces deux tendances complémentaires car elles contribuent à une définition plus utile de la notion de validité qui a plusieurs types : la validité prédictive, concourante, de contenu, de critère... Dans ce sens, le « *Manuel pour l'Elaboration et la Passation de Tests et d'Examens de Langue* » avance que ce critère qualitatif « *dépend de chaque étape du cycle d'élaboration et de passation du test. La validité se construit tout au long de l'ensemble du processus* » (2011, p. 17).

En dépit de son importance, ce critère de validité n'est pas toujours respecté lors de la conception des examens. Dans ce contexte, Romainville affirme que « *les outils d'évaluation mesurent parfois des compétences qui n'ont pas fait l'objet d'un enseignement* » (2011, p. 3), d'où les jugements subjectifs. Ce qui nous amène à souligner la nécessité de la validation des examens avant leur passation, une validation qui assure la vérification de la capacité des items sélectionnés à évaluer les compétences visées.

Nous ne nous attardons pas sur ces procédures de validation évoquées par de nombreux auteurs (De Landsheere, 2000 ; De Ketele & Gérard, 2005...) car elles seront abordées dans le chapitre consacré à la validité de l'évaluation.

#### **4.2. La fidélité (fiabilité)**

Selon Mothe, le critère de fidélité est moins important que celui de la validité dans la mesure où retenir « *un instrument qui mesure sans précision ce qu'on veut mesurer est préférable à un instrument qui mesurerait avec précision...autre chose* » (1975, p. 15). Par opposition à la validité qui est déterminée par la conformité du contenu du test à ses objectifs, la fiabilité se manifeste par la constance des résultats des évalués dans des tests passés dans les mêmes conditions. Si « *les résultats obtenus sont aussi exempts que possible d'erreurs de mesure* » (Noel-Jothy & Sampsonis, 2006, p. 42), le test sera fidèle. Mentionnons que la constance des résultats ne dépend pas de la constance des instruments de mesure retenus. Lorsque le test est fidèle, selon De Landsheere, « *faire subir la même épreuve à diverses reprises ne permet pas un bon contrôle de la fidélité* » (2000, p. 194).

Pour De Landsheere (2000), le contrôle de la fidélité des examens s'effectue à deux niveaux : les questions du test et la notation. La fiabilité du test pourrait être garantie à travers :



la clarté et la précision des questions proposées, le nombre d'items... Pour le premier élément, les questionnaires et les situations proposées en évaluation ne doivent pas être ambiguës. Ainsi, la clarté et la précision des consignes assurent des réponses fiables. Dans cet ordre d'idées, Noel-Jothy et Sampsonis (2006) reprennent des conseils pour la passation et la correction de tests, extraits du « *Guide des correcteurs/examineurs du CEFPI niveau A2* ». Parmi ces conseils, nous citons la liberté donnée à l'examineur qui « *ne doit pas hésiter à reformuler la consigne, voire l'expliquer au candidat dans sa langue maternelle, s'il voit que celui-ci ne comprend pas* » (Noel-Jothy & Sampsonis, 2006, p. 54). En éclaircissant les consignes par explication, les réponses des candidats seront plus précises et fiables.

Le nombre d'items que comporte un examen est aussi un élément à prendre en considération pour la conception de tests fiables. Ainsi, selon De Landsheere, « *plus on ajoute de questions, plus la fidélité augmente* » (2000, p. 195). Sans tenir compte des contraintes suscitées par un examen long, cette citation associe la fidélité de tests au nombre de questions qu'ils incluent. Les questions nombreuses couvriraient tout le contenu de la matière alors que peu de questions en négligent des parties. Pour cette raison, il est important de sélectionner des items représentatifs du contenu enseigné par échantillonnage (De Landsheere, 2000). Ce deuxième élément est également lié à la validité des tests qui est mesurée par la représentativité des items retenus.

En plus de ces éléments relatifs à la fiabilité des réponses, la fiabilité de la notation est garantie en éliminant les biais de la notation (effets de contamination, d'ordre, de halo...). Soucieux de notes fiables, certains enseignants procèdent à deux corrections séparées par une durée de temps. D'autres optent, par concertation, à la correction d'une même copie par différents enseignants. Le rôle que joue l'accord des enseignants au niveau de la fiabilité des notes est d'une importance considérable étant donné qu'ils doivent appliquer les mêmes critères et suivre le même barème (Piccardo & al, 2011).

Avant de poursuivre les autres critères professionnels de l'évaluation, nous tenons à préciser que des liens sont établis entre les deux critères professionnels cités ci-dessus : validité et fiabilité. Ainsi, la fidélité est présentée comme une forme de validité (Cureton, 1951 ; André, Loye & Laurencelle, 2015) ou une condition de celle-ci (Hotyat, 1962 ; Rolland, 2006). Par opposition à ces deux positions, De Landsheere (2000) et Mothe (1975) affirment que la fidélité du test n'en assure pas la validité étant donné que la constance, la stabilité et l'objectivité des résultats ne signifient pas que la compétence visée est réellement mesurée.

### 4.3. L'objectivité

Dimension de la fiabilité (Mothe, 1975 ; Noel-Jothy & Sampsonis, 2006), l'objectivité constitue le troisième critère professionnel de l'évaluation. Elle consiste en « *une qualité d'un test qui fait que les résultats obtenus par un sujet donné seront constants quel que soit l'examineur* » (Noel-Jothy & Sampsonis, 2006, p. 42). Il est nécessaire que les corrections des réponses données par un apprenant évalué, effectuées par des enseignants variés, aboutissent à une note identique, pour une évaluation objective. Opposée à l'objectivité, la subjectivité se manifeste à trois niveaux du test évaluatif : la conception des items, les réponses des apprenants et leur correction (Pilliner, 1968). Cependant, « *la distinction entre un examen subjectif et un examen objectif réside donc uniquement dans la manière dont les notes sont attribuées* » (Mothe, 1975, p. 7). Etant donné que l'objectivité de l'évaluation est exclusivement liée à la correction, le CECRL (2011) introduit la notion de *correction objectivée*.

Puisque l'objectivité en évaluation est une dimension de la fidélité et de la validité, elle devient le souci de nombreux acteurs concernés par cet acte (Gérard, 2002). Cependant, ceux-ci ne sont pas en mesure d'éradiquer entièrement la subjectivité, une élimination qui requiert « *des évaluateurs délibérés de leur condition humaine* » (Erard, 2015, p. 14). Dans ce sens, de nombreux auteurs affirment que l'évaluation objective est une chimère (Bonniol & Vial, 2009 ; Romainville, 2011) dans la mesure où de nombreux effets en biaisent les résultats (effet de Halo, de contraste, d'assimilation...). A ces effets incontestables, s'ajoutent des raisons relatives à la pratique enseignante. Ainsi, la variété des critères de correction et la valeur que les correcteurs leur accordent, en cas du recours à des critères communs, sont des facteurs qui nuisent considérablement à l'objectivité de la notation et de l'évaluation (Romainville, 2011).

Romainville estime que la subjectivité en évaluation, certes inévitable, pourrait être objectivée par diverses stratégies. Selon lui, « *objectiver la subjectivité revient à tout à la fois à reconnaître aux « sujets pensants » évaluateurs la légitimité d'un point de vue spécifique, mais à réclamer aussi que ce point de vue ait été l'objet d'une explicitation préalable auprès des étudiants* » (2011, p. 6). Ainsi, les enseignants minimisent les effets de leur jugement subjectif à travers la communication antérieure des objectifs, des critères et des modalités de l'évaluation, le développement de l'intersubjectivité (confrontation des points de vue évaluatifs lors des corrections collectives)... (Romainville, 2011).

### 4.4. La faisabilité

Tout comme l'objectivité, le critère de faisabilité ou praticabilité est également lié aux qualités fondamentales de toute évaluation : validité et fiabilité (Lussier & Turner, 1995). La

faisabilité se rapporte, selon Noel-Jothy et Sampsonis, à la « *facilité de réalisation à tous les niveaux. La clarté des instructions, la présence d'exemples, la durée impartie à la tâche par rapport à sa longueur ou sa difficulté font partie de ce critère* » (Noel-Jothy & Sampsonis, 2006, p. 42).

La clarté des consignes est une condition de fiabilité, de validité et de praticabilité des tests. Pour collecter les réponses des apprenants évalués destinées à être corrigées, notées et interprétées, les examinateurs ont à sélectionner des consignes précises et éviter les items ambigus, ce qui assure des informations utiles sur la compétence évaluée. En plus de ce premier élément qui garantit la faisabilité des examens, la durée consacrée à la passation est à prendre en considération lors de la conception des tests. En effet, les examinateurs sélectionnent des activités ou des situations dont la longueur et la complexité dépendent de ce second indicateur. Un examen ne sera pas faisable si les concepteurs proposent de nombreuses questions auxquelles les évalués doivent répondre en peu de temps. Dans « *Manuel pour l'Elaboration et la Passation de Tests et d'Examens de Langue* », la durée impartie au test est délimitée de façon à ce que le candidat moyen puisse « *disposer d'assez de temps pour répondre à tous les items du test ou de l'examen sans avoir à se presser. L'essentiel est que les candidats aient l'occasion de montrer leur capacité réelle* » (2011, p. 24). A ce niveau, nous constatons que de nombreux enseignants-examinateurs adaptent les contenus (le choix des tests, le nombre des questions posées, la complexité des situations de rédaction...) de leurs tests à la contrainte temporelle (Piccardo & al, 2011).

Les deux critères de faisabilité et d'objectivité, en dépit de leur utilité, ne figurent pas parmi les critères listés dans « *Manuel pour l'Elaboration et la Passation de Tests et d'Examens de Langue* ». Cette omission des deux qualités est certainement due à leur imbrication dans les qualités globales : la validité et la fiabilité.

#### **4.5. L'équité**

Dernier critère cité par Noel-Jothy et Sampsonis (2006), il renvoie à l'absence de biais qui altèrent une évaluation valide des compétences effectives des évalués. En effet, la prise en considération des aspects qui ne se rapportent pas aux dites compétences ne conduit pas aux décisions appropriées (Erard, 2015). Pour désigner le traitement équitable des apprenants en évaluation, assurant une évaluation valide, Mègre et Riba (2014) introduisent le critère de « *standardisation* ». Ce critère se réalise « *en harmonisant, dans le temps et dans l'espace, toutes les procédures liées à la conception des épreuves, à leur passation, ainsi qu'à la*

*correction des productions des candidats et de leurs feuilles de réponses* » (Mègre & Riba, 2014, p. 64.).

Cette harmonisation, qui est certainement nécessaire pour une égalité de chances, ne pourrait pas être entièrement équitable vu la diversité physique, intellectuelle et culturelle des évalués (Jeffrey, 2013), d'où les adaptations introduites par les enseignants. Ces adaptations correspondent à une conception de l'équité, préconisée par Allal et Mottier Lopez qui avancent : *« plutôt que d'avoir une idée d'équité formelle selon laquelle chaque élève est traité de manière identique, les enseignants envisagent l'équité comme le traitement le plus adéquat possible pour assurer un avenir scolaire de qualité à chaque élève »* (2008, p. 231). Comment les particularités des apprenants pourront-elles être prises en considération ?

Dans *« Manuel pour l'Elaboration et la Passation de Tests et d'Examens de Langue »*, le choix des sujets par les examinateurs ou les concepteurs constitue une manière de minimiser les biais dans la mesure où *« certains sujets (par exemple les coutumes locales) peuvent avantager ou désavantager certains groupes de candidats (par exemple ceux qui viennent de pays où les coutumes sont très différentes) »* (2011, p. 19). Cette citation aborde le critère de l'équité au niveau de la conception des tests. Ce niveau-ci peut être intégré dans les aspects de ce critère mentionnés par *« Les Standards »* (1999) : réduction de biais, traitement équitable en évaluation et égalité de correction.

Les critères que nous venons de lister en nous inspirant de Noel-Jothy et Sampsonis (2006) ne sont pas exhaustifs. Cependant, ce sont des critères reconnus par de nombreux chercheurs (Mothe, 1975, De Landsheere, 2000, Mègre & Riba, 2014...). Le *« Manuel pour l'Elaboration et la Passation de Tests et d'Examens de Langue »* mentionne le souci *éthique* des examinateurs lors de la conception, la passation et la correction du test. Ceux-ci garantissent une pratique évaluative éthique à travers une analyse réflexive (Gautier, 2005 ; Lafortune, 2008).

#### **4.6. La validité : pour une évaluation authentique**

De tous les critères expliqués supra, celui de la validité est certainement le plus important dans les divers dispositifs d'évaluation. Cette affirmation de Mothe (1975), que nous admettons dans notre étude, est peu conforme à la conception des auteurs qui opposent la validité aux autres critères (fidélité, objectivité...). De nombreux auteurs (Kelley, 1927 ; Lado, 1961 ; Mothe, 1975 ; De Landsheere, 2000, Riba & Mègre, 2014...) sont unanimes sur la définition de la validité en tant que la capacité de tests, épreuves et examens à mesurer ce qui

est mesuré ou évalué. Cependant, cette définition est expliquée de deux manières distinctes, qui nous interpellent à ce niveau.

Du premier point de vue, la validité consiste au recours à une tâche évaluative appropriée pour évaluer une compétence précise, comme le mentionnent Riba et Mègre (2014). Les propos de ces auteurs sont illustrés par l'exemple suivant : « *un candidat qui répond à un questionnaire de compréhension écrite ne peut être évalué, à partir de ce même questionnaire, en production écrite si ses réponses comportent des erreurs grammaticales ou orthographiques* » (Riba & Mègre, 2014, p. 63). Dans le même sens, Lado précise que « *si un test de prononciation mesure la prononciation et rien d'autre, c'est un test valide de prononciation, ce ne serait pas un test valide de grammaire ou de vocabulaire* » (1961, p. 30). Dans ces deux illustrations, la validité est prise comme une qualité particulière du test qui exige que celui-ci mesure une compétence précise. Ainsi, dans le cadre de l'évaluation de la compréhension écrite, la formulation des consignes et des questions de compréhension proposées doit être précise et claire (Beaume, 1988) pour évaluer seulement ce qui est évalué (la compréhension du texte) et ne pas sanctionner ce qui ne l'est pas (la non-compréhension des questions). Ce premier point de vue privilégie la validité à la fidélité car évaluer la compétence à évaluer est préférable à évaluer avec précision une autre compétence (Mothe, 1975).

Le second point de vue, contrairement au premier, présente la validité comme une qualité générale : celle-ci englobe les autres critères que nous venons de lister (fidélité, objectivité, équité...). La raison de cet aspect général de la validité est évidente : une évaluation valide de telle ou telle compétence requiert également la précision des instruments de mesure (fidélité), une objectivité lors de la correction des copies... Riba et Mègre (2015) ont illustré le critère de validité par le questionnaire de compréhension écrite qui ne doit pas évaluer des compétences autres que cette compréhension. Cependant, pour que ce questionnaire soit valide, n'est-il pas nécessaire qu'il évalue avec précision et objectivité cette compétence visée ? Peut-on considérer valide une évaluation qui ne mesure pas réellement ce qui est évalué ? Certes, une évaluation valide est une évaluation authentique qui englobe tous les critères professionnels de l'évaluation sommative. En effet, la validité intègre toutes les phases du processus d'évaluation, d'où son caractère général.

L'importance de ce critère et son caractère global expliquent le choix de notre thème de recherche qui porte sur la validité des dispositifs évaluatifs dans le système éducatif algérien. Nous aborderons ce critère de manière plus détaillée dans le chapitre suivant qui sera consacré

à ce critère de validité pour mieux définir la notion et en tracer les contours. Étant concises, les informations fournies supra pourraient omettre des données indispensables à notre étude.

### **5. Fonctions de l'évaluation sommative**

Bélaïr (1999) et Perrenoud (2001) ont listé plusieurs fonctions de l'évaluation de manière détaillée. Le premier cite les trois fonctions diagnostique, formative et sommative, alors que le second liste les fonctions : pronostique, formative et certificative. De ces fonctions, nous énumérons celles de l'évaluation sommative retenues dans le travail de Bélaïr (1999).

Dans son travail, les fonctions sont associées aux différents types d'évaluation. Ainsi, chaque type a de nombreuses fonctions : certaines correspondent aux enseignants alors que d'autres sont liées aux élèves. Pour l'évaluation sommative qui intervient au bout des apprentissages, elle permet aux enseignants de :

- *Vérifier* les connaissances et les compétences des apprenants au bout des enseignements dispensés.
- *Valider* les compétences réellement acquises, mais aussi les compétences insuffisamment ou non acquises.
- *Inform*er les divers acteurs impliqués des compétences des apprenants.
- *Orienter* les évalués vers des disciplines qui correspondent à leurs aptitudes.

Quant aux apprenants, cette évaluation contribue à :

- Choisir conformément à ses aptitudes (une spécialité, une discipline, un métier...).
- Entreprendre des formations pour maîtriser ce qui ne l'est pas.

Contrairement aux autres types évaluatifs, dont les retombées ne franchissent pas l'univers scolaire, les résultats de l'évaluation sommative sont destinés aux parents des élèves. L'évaluation devient ainsi un outil informatif qui leur fournit des renseignements sur les compétences acquises par leurs enfants.

### **6. Évaluation sommative : une forme souvent contestée**

En dépit de ses fonctions, l'évaluation sommative continue à être une forme primordiale dans le parcours de tout apprenant dans la mesure où ses résultats aboutissent à des décisions importantes : la réussite ou l'échec, l'orientation... Néanmoins, certaines critiques lui sont assignées par des auteurs comme Viallet et Maisonneuve (1981), Petitjean (1984), Prégent (1990) et Vidaillet (2013).

Selon Petitjean (1984), quelques aspects de l'évaluation sommative sont à l'origine des multiples reproches qui lui sont faits. Le premier aspect est le caractère global de cette évaluation, qui évalue dès le début des apprentissages un ensemble d'objectifs. Cette première

critique ne nous semble pas fondée dans la mesure où une autre forme au caractère détaillé existe : l'évaluation formative. Ces deux formes aux fonctions particulières sont complémentaires : la forme formative détaillée prépare les apprenants à la forme sommative qui porte sur la totalité des contenus enseignés. Le deuxième aspect présenté par cet auteur se rapporte à l'objet évalué. Contrairement à la forme formative, la forme sommative de l'évaluation ne tient pas compte des démarches adoptées par les apprenants pour la résolution de situations évaluatives sélectionnées. De cette manière, cette évaluation fournira peu de renseignements aux enseignants sur les ressources, les processus et les stratégies mobilisés par leurs apprenants pour la résolution des tâches proposées. Le dernier aspect que nous reprenons dans notre étude est le caractère normatif de l'évaluation sommative qui vise le classement des apprenants et la comparaison de leurs performances. Cette comparaison ne devrait pas être visée dans une évaluation intelligente. Cet aspect est également difficile à admettre vu qu'une approche critériée est également possible en évaluation sommative.

Petitjean, après avoir listé ces *reproches*, reprend *les effets pervers* de cette évaluation sommative que citent Viallet et Maisonneuve (1981). Ces effets, nous les résumons dans les points suivants :

- Validité des instruments d'évaluation

Ces instruments concernent en premier lieu les examens et les tests que les enseignants proposent et qui ne sont pas suffisamment valides pour représenter avec exactitude les compétences des apprenants. Si ces tests révèlent certaines connaissances et compétences des apprenants, ils n'en assurent pas la durabilité ou le transfert dans des situations scolaires et authentiques nouvelles. En second lieu, les instruments abordés incluent les outils de jugement investis pour représenter les résultats de l'évaluation : la note. Celle-ci détermine les décisions prises par le personnel enseignant (remédiation, sanction...), écartant ainsi le rôle que doit jouer un aperçu réel sur les compétences des apprenants dans la prise de décisions correctes.

- Conditions psychologiques défavorables

Selon les mêmes auteurs, l'évaluation sommative dans le milieu scolaire génère des conditions défavorables à tout apprentissage linguistique. En effet, cette forme évaluative installe une certaine insécurité chez les apprenants qui se sentent menacés par les procédures d'évaluation. Une insécurité pareille les pousse à s'intéresser à la nature des activités à proposer en évaluation au détriment du développement de leurs compétences. De cette manière, l'évaluation sommative est à l'origine d'une attitude complètement indésirable dans l'enseignement/apprentissage de toute discipline. Ce sentiment d'insécurité s'intensifie en

raison de la succession d'évaluations sommatives au cours de l'année scolaire, notamment que leur nombre pourrait dépasser les exigences.

En plus de l'insécurité créée par cette évaluation, des auteurs comme Ames (1992), Covington (1992) et Viau (2002) dénoncent la démotivation résultant de cette évaluation sommative chez une catégorie d'apprenants. En effet, Viau reprend la réflexion de Covington qui précise que « *les pratiques évaluatives centrées sur la performance suscitent parfois la motivation des plus forts, mais nuisent généralement à celle des élèves moyens et faibles* » (Viau, 2002, p. 77). Contrairement aux bons résultats qui motivent les apprenants qui les obtiennent, les mauvais résultats découragent souvent les apprenants faibles et moyens qui se sentent incapables de progresser. La note à laquelle cette évaluation aboutit est ainsi perçue comme une source de démotivation, occultant une condition clé à tout apprentissage, ce qui explique la position de certains auteurs qui continuent à plaider pour sa suppression en tant qu'outil de jugement. Ce ne sont pas cependant tous les auteurs qui rejoignent cette conception étant donné que c'est la motivation intrinsèque qui assure un réel engagement cognitif et affectif, comme le précise Viau (2002).

- Vers un enseignant-évaluateur

L'évaluation sommative requiert des enseignants une préparation régulière des examens et l'organisation des conditions de leur passation. Une telle situation contraint les enseignants à chercher des supports et des activités nouvelles pour chaque évaluation. Le temps consacré à cette préparation consomme celui de la préparation des cours et le suivi des progrès atteints par les apprenants. De cette manière, l'enseignant se transforme en évaluateur qui contrôle périodiquement les connaissances et les acquis de ses apprenants, négligeant sa fonction principale qui consiste au développement de leurs compétences. En vue d'illustrer la domination de la fonction évaluative sur la fonction pédagogique en classe, Prigent attribue à l'enseignant le rôle de *censeur* dans la citation suivante où il avance que « *le professeur joue (...) le rôle d'un censeur qui détermine, à l'aide d'une épreuve scolaire, si la somme des connaissances acquises par un étudiant au terme d'une étape du cours est suffisante* » (1990, s.p.<sup>23</sup>).

- Carence informative

---

<sup>23</sup> Page non indiquée (Google Books).



En plus des effets pervers de la forme sommative listés, les auteurs soulignent le manque de renseignements apportés sur les stratégies adoptées par les apprenants et leurs lacunes, ce qui élimine incontestablement les interventions enseignantes efficaces. Cette réflexion rejoint celle de Petitjean citée supra : l'évaluation sommative ne fournit pas des renseignements utiles au progrès des apprenants et au développement de leurs compétences. La position de ces auteurs rehausse la valeur de l'évaluation formative qui fournit des indications extrêmement utiles sur les démarches empruntées par les apprenants pour résoudre les problèmes proposés, les ressources qu'ils investissent, les stratégies auxquelles ils recourent... Cette précision des renseignements est garante d'une intervention enseignante adéquate. Cet effet pourrait être contesté en raison du croisement des fonctions des diverses formes d'évaluation : la fonction régulatrice –autrefois exclusive à la forme formative- est actuellement admise en évaluation sommative.

Ces effets que nous venons de reprendre en nous appuyant sur les travaux de Viallet et Maisonneuve (1981) sont également cités par Prégent dans son ouvrage « *La préparation d'un cours* ». Celui-ci évoque les contraintes de l'évaluation sommative du point de vue des enseignants et de celui des apprenants. Nous reprenons ces contraintes dans les éléments suivants de manière concise étant donné qu'elles correspondent aux effets pervers tels que les auteurs antérieurement cités les évoquent :

- Du point de vue des enseignants

Pour les enseignants, qui ont parfois la possibilité de déterminer le nombre d'examens à intégrer (Prégent, 1990), l'évaluation sommative les contraint à élaborer périodiquement des sujets d'examens et à corriger les réponses de leurs apprenants. La conception de ces examens est une tâche qui exige des efforts, ainsi que la collaboration des enseignants chargés de la matière. Ceux-ci doivent sélectionner des questions originales qui constituent un échantillon représentatif des objectifs et des contenus enseignés. Il en est de même pour la correction des copies qui ne pourrait pas être exempte des effets qui nuisent à la validité des notes attribuées aux apprenants. Cette situation qui donne à l'évaluation une priorité en classe conduit les enseignants à négliger les pratiques enseignantes au profit des pratiques évaluatives. Ces enseignants ont même tendance à conformer leurs pratiques enseignantes aux activités évaluatives qui seront proposées en évaluation sommative. De ce fait, la préparation des apprenants aux examens prime sur le développement de compétences réelles chez eux, ce qui explique clairement les résultats négatifs que présente une évaluation sommative.

- Du point de vue des apprenants

La forme sommative détourne l'attention des apprenants du suivi des cours et des contenus enseignés. Ceux-ci, qui pourraient être démotivés par des résultats insatisfaisants en évaluation, centrent leur intérêt sur les questions que leurs enseignants pourraient proposer (Prégent, 1990).

Bien qu'elles soient globales, ces contraintes que nous venons de citer en nous référant aux travaux de Viallet et Maisonneuve (1981), Petitjean (1984) et Prégent (1990) ne sont pas exhaustives, car d'autres critiques ont été formulées par des auteurs que nous n'avons pas encore cités. Dans ce contexte, les critiques adressées par Jorro (2000) à cette forme évaluative illustrent nos propos. En effet, cet auteur insiste sur les problèmes de validité et de fiabilité de la notation, outil investi par les enseignants. Cet élément est contraignant aux enseignants qui doivent recourir à des instruments évaluatifs valides et créer des conditions de correction favorables à une correction valide. De telles conditions contribuent à réduire les écarts qui surgissent entre *le jugement réel*<sup>24</sup> et *le jugement formulé* par les enseignants en raison de plusieurs effets comme la fatigue, la comparaison des copies... Nous ne développons pas –à ce niveau- ces effets qui altèrent la notation et la validité des notes car nous les évoquerons ultérieurement.

## 7. Des pratiques qui relèvent de l'évaluation sommative

Dans l'univers scolaire, l'évaluation sommative se réalise à travers des pratiques variées. Par *pratiques*, nous ne désignons pas les activités particulières à l'évaluation des langues, mais les évaluations introduites par les enseignants au terme des apprentissages pour établir un bilan des acquis. Il s'agit des épreuves, examens, tests, travaux, exercices ... introduits au terme d'un ensemble d'enseignements. Nous ne pouvons à ce niveau nous étaler sur les activités auxquelles les enseignants recourent en évaluation, un élément qui sera traité dans notre partie pratique à travers l'analyse des évaluations sommatives collectées. Nous préférons définir les *termes* les plus fréquents, désignant les pratiques qui relèvent de l'évaluation sommative dans le milieu scolaire, en vue de déterminer déceler les différences existantes entre ces termes qui pourraient être confondus. Vu la confusion terminologique mentionnée entre les termes test et examen, souvent employés comme synonymes comme le précise Morissette (1993), nous définissons les deux concepts de manière détaillée tout en

---

<sup>24</sup> L'assemblage de ces deux mots semble exagéré vu la contradiction de leurs sens : le mot *jugement* sous-entend un avis personnel et subjectif sur une question précise, et devient peu compatible avec le mot *réel* qui représente ce qui est prouvé et dépourvu de subjectivité. Nous l'employons pour désigner un jugement qui décrit le mieux la compétence évaluée, c'est-à-dire qui reflète la réalité objectivement.

précisant quel terme nous devons maintenir pour désigner les évaluations sommatives proposées aux apprenants.

### 7.1. Les examens

Dans son ouvrage « *Les examens : Les moyens d'Évaluation dans l'Enseignement* », Hoytat cite la conception classique du terme *examen*. Ainsi, Buisson considère l'examen comme « *une sanction officielle, indispensable pour marquer le but et pour obliger la jeunesse à y tendre un effort plus énergique et plus soutenu* » (1911<sup>25</sup>, cité dans Hoytat, 1962, p. 15). Cette définition, qui dans sa première partie réduit ce moyen d'évaluation à une sanction ou punition, ne néglige pas une des fonctions actuelles de ces examens : la fonction formative. En effet, contrairement à la conception classique, la conception récente accorde des fonctions multiples aux examens, impliquant tous les acteurs pédagogiques et institutionnels. Nous reprenons brièvement ces fonctions, en nous référant à Hoytat.

Pour les enseignants, les examens ont une fonction diagnostique (contrôler les prérequis nécessaires à tout nouvel enseignement), une fonction formative (fournir des renseignements utiles à la remédiation aux insuffisances) et une fonction sommative (vérifier les apprentissages, classer les apprenants, les orienter...). Quant aux apprenants, comme la définition de Buisson le suggère, les examens les secouent pour prodiguer des efforts et améliorer leurs performances. Néanmoins, dans certains cas, les résultats obtenus produisent des effets inverses : au lieu de les motiver, ils les découragent. C'est dans cette perspective que certains enseignants perçoivent toujours les examens comme générateurs de conditions psychologiques défavorables aux apprentissages et que les apprenants les appréhendent comme un mal inévitable. En plus de ces deux acteurs pédagogiques, les résultats de l'évaluation sommative sont utiles aux parents qui doivent être continuellement informés du profil éducatif (comportement, discipline...) et du profil cognitif (degré de maîtrise des connaissances et compétences visées). En informant les parents, les enseignants sollicitent leur contribution dans la formation personnelle et scientifique de leurs enfants car le milieu familial est également important au développement des compétences de leurs enfants. Enfin, nous devons mentionner l'impact des examens sur le plan institutionnel et social étant donné qu'ils constituent des moyens pour évaluer les enseignants et les programmes, qu'ils lèguent aux établissements scolaires une réputation et une crédibilité aux yeux des parents, des autres établissements et de la société...

---

<sup>25</sup> Nous n'avons pas cité la page car nous n'avons pu accéder qu'à la version électronique de ce dictionnaire accessible sur le site <http://www.inrp.fr/edition-electronique/lodel/dictionnaire-ferdinand-buisson/>

Ces fonctions que l'auteur cite ne sont pas cependant particulières aux examens et ne fondent pas les particularités de la notion. Mais, à travers la typologie des examens établie par l'auteur, qui distingue : les examens à l'échelle régionale ou nationale, les épreuves classiques ou objectives, les examens prédictifs ou de fin d'études<sup>26</sup>..., nous pouvons dégager le caractère global des examens qui sont destinés à un public restreint ou large, constitués d'activités variées, intégrés à tous les moments du processus pédagogique... Trouvant cette perspective définitoire adoptée par l'auteur peu utile, nous nous référons à l'auteure Morissette qui, en collaboration avec Laurencelle, recourent à un critère qui permet de distinguer les examens. Ces auteurs les conçoivent en effet comme des épreuves qui ne sont pas standardisées ou qui le sont peu. Mais, que signifie cette standardisation ?

Pour expliciter ce que désigne la standardisation, nous reprenons ses quatre conditions listées par ces auteurs :

1. Tous les étudiants répondent aux mêmes questions.
2. Tous les étudiants reçoivent les mêmes instructions claires et précises.
3. Aucun étudiant ne profite d'avantages externes à l'objet de mesure par rapport aux autres.
4. Pour tous les étudiants, un système prévu de correction s'applique uniformément (Morissette & Laurencelle, 1993, p. 69)

Ces conditions qui assurent la standardisation des évaluations deviennent absentes dans les examens ou peu présentes comme l'avancent les deux auteurs dans leur citation. Une telle réalité existe dans les évaluations courantes dans le milieu scolaire où, dans certains cas, les enseignants sont contraints de proposer aux apprenants d'un même niveau des questions différentes selon leur état d'avancement dans les cours retenus. Dans certains cas, et en raison de contraintes institutionnelles, les apprenants ne passent pas ces examens dans les mêmes conditions, ce qui aurait des retombées sur les instructions reçues, les explications fournies... Enfin, les enseignants peuvent adopter des outils correctifs variés en raison de contraintes particulières. Les examens ne sont pas uniformes que dans certains cas, d'où la définition que les mêmes auteurs proposent dans le même ouvrage. Selon eux, l'examen consiste en :

Une épreuve ou une série d'épreuves, aussi uniformes que les circonstances le permettent, grâce auxquelles obtient des résultats, la plupart du temps quantitatifs, parfois qualitatifs,

---

<sup>26</sup> Ces types d'examens, distingués en fonction de critères bien précis : organisation, contenu, fonction, sont expliqués brièvement par F. Hotyat (1962).

soit sur l'atteinte d'un certain nombre d'objectifs, soit sur la performance en rapport avec un contenu scolaire (Morissette & Laurencelle, 1993, p. 66).

Cette définition, qui introduit le mot épreuve comme la définition citée supra, suscite des interrogations sur le lien entre les termes examens et épreuves notamment que le terme épreuves est également employé pour désigner les évaluations proposées dans le milieu scolaire. Cette confusion, que nous évoquerons dans les éléments suivants, ne doit pas détourner notre attention sur les caractéristiques des examens que la citation met en avant. La première caractéristique est liée à *la standardisation* des épreuves qui, contrairement aux tests, ne constituent pas une condition dans les examens. De préférence *uniformes*, les examens peuvent ne pas l'être si les circonstances ne le permettent pas. La deuxième caractéristique qui interpelle notre attention est la nature des résultats auxquels les épreuves aboutissent. Souvent quantitatifs, ces résultats mesurent essentiellement le degré de maîtrise des objectifs ciblés par les enseignants. Néanmoins, ils peuvent être qualitatifs si les enseignants évaluent les compétences visées par le programme, à travers une évaluation des performances. Ces informations fournies par les définitions proposées par les deux auteurs cités ci-dessus nous semblent très utiles pour définir la notion.

Si les définitions précédentes insistent sur les fonctions et les caractéristiques des examens, la définition que propose Belhoste met en exergue les moments de ces modes évaluatifs. Il estime que, dans les examens scolaires, « *on retrouve (...) deux opérations successives de l'épreuve et du jugement* » (2008, p. 2) : un examen comprend une ou plusieurs épreuves qui sont à l'origine de décisions relatives aux apprenants (réussite ou échec), ou *verdicts*<sup>27</sup>. Ces deux opérations sont prises en considération par cet auteur dans sa classification d'examens. En effet, celui-ci cite des types d'épreuves (épreuves orales ou écrites) et des types de verdict (examens de passage ou examens de classement). Cette typologie de verdict nous semble liée aux fonctions des examens que nous distinguons des types. En retenant la première opération citée par Belhoste (2008), nous trouvons que cet auteur ainsi que Morissette et Laurencelle sont unanimes sur la définition des examens en tant qu'épreuves. Soulignons que les examens sont, selon les derniers auteurs, le mode évaluatif le plus pratiqué par les enseignants dans le milieu scolaire.

---

<sup>27</sup> Concept employé par Belhoste qui estime que les examens comprennent deux opérations : épreuves et verdicts c'est-à-dire décisions.

Contrairement à Morissette et Laurencelle qui déterminent le critère de standardisation comme un trait distinctif entre test et examen, Riba et Mègre l'intègrent parmi les critères professionnels de tout outil ou instrument évaluatif. Ces deux auteurs introduisent un autre trait distinctif : l'obtention de diplôme. Ils estiment que tout examen « *conduit, dans la plupart des cas, à un diplôme valable pour toute la vie (baccalauréat [...])* » (Riba & Mègre, 2014, p. 91). Bien que ces auteurs reconnaissent le caractère sommatif des examens, ils les situent seulement au bout d'un cycle complet d'apprentissages. Cette restriction nous conduit à nous interroger sur le terme que ces auteurs emploieraient pour désigner les évaluations sommatives introduites au bout d'une séquence, d'un trimestre ou d'un semestre. En effet, des évaluations sommatives sont introduites au bout d'une série d'apprentissage dont la durée ne pourrait être délimitée avec précision étant donné qu'elle relève souvent de choix institutionnels.

En nous référant essentiellement à l'ouvrage de Hotyat « *Les examens : Les moyens d'évaluation dans l'enseignement* », nous abordons dans ce qui suit les épreuves que regroupent les examens scolaires. En effet, l'auteur liste les épreuves écrites, les rédactions et les examens oraux. Ces épreuves, nous les expliquerons dans les éléments suivants :

### **7.1.1. Les épreuves écrites**

Selon Hotyat (1962), ces épreuves consistent en des activités de reproduction, application, critique, analyse ou de synthèse. Les apprenants devraient dans la première catégorie reproduire des parties des cours dispensés par leurs enseignants qui évaluent dans ce cas-ci leurs connaissances. La deuxième catégorie requiert l'application des connaissances acquises dans des situations nouvelles. C'est le savoir-faire des apprenants qui est évalué à ce niveau, contrairement à la première catégorie d'activités qui évaluent le savoir déclaratif des apprenants. Les activités d'analyse qui consistent à établir des liens entre des éléments étudiés sont également intégrées dans les épreuves écrites, ainsi que les activités de synthèse. Dans cette dernière catégorie, les apprenants investissent les savoirs cumulés dans les cours appris pour constituer un tout cohérent et harmonieux. En considérant ces catégories d'activités, nous trouvons qu'elles correspondent aux diverses catégories de la taxonomie des objectifs cognitifs de Bloom : connaissance, compréhension, application, analyse, synthèse et évaluation. En effet, ces niveaux taxonomiques ont depuis longtemps constitué une référence pour la conception des cours et des examens : les enseignants s'y réfèrent pour préparer leurs leçons en fonction d'objectifs cognitifs précis qui seront évalués dans le cadre des examens organisés au bout des apprentissages programmés (Morissette, 1993).

Le privilège accordé à ces épreuves écrites est dû selon Hotyat à des raisons psychologiques et méthodologiques. Nous les regroupons dans les points suivants :

- La précision des objectifs guide les apprenants qui auront à apprendre des contenus précis, comparer leurs performances à celles attendues dans le programme...
- La communication écrite des connaissances est plus difficile, d'où la nécessité d'évaluer cette compétence chez les apprenants.
- Le contrôle efficace des connaissances est assuré par des activités écrites qui évaluent la capacité des apprenants à mobiliser ce qui est appris dans des contextes appropriés.

En dépit de leur abondance comme instruments d'évaluation, de nombreuses critiques lui sont, à raison, assignées. Parmi ces critiques, citons certaines listées par Hotyat : le nombre restreint de questions dans une épreuve qui entrave une évaluation efficace des contenus retenus, la nature de certaines questions regroupant plusieurs objectifs qui offrent peu de renseignements sur les progrès des apprenants et leurs lacunes, la fidélité des notes qui ne pourrait être assurée vu la diversité des correcteurs... Ces critiques, aussi importantes soient-elles, n'ont pas amoindri l'usage de ces épreuves dans les univers scolaires. En effet, ces épreuves sont toujours introduites pour évaluer les acquisitions des apprenants notamment que tous les niveaux taxonomiques y sont inclus.

### **7.1.2. Les rédactions**

La deuxième forme d'examen, traditionnelle selon Hotyat, est la rédaction. Cette activité requiert des apprenants évalués la composition de textes sur un sujet précis imposé. Fréquemment, les enseignants laissent à leurs apprenants la possibilité de choisir un thème parmi une série de sujets proposés. Dans les deux cas, des critiques véhémentes lui ont été attribuées et ce, pour de multiples raisons que nous reproduisons –en nous référant à Hotyat– dans ce qui suit :

- Un sujet imposé ne relève pas nécessairement des intérêts des apprenants qui pourraient avoir très peu de renseignements sur celui-ci. De ce fait, leurs rédactions sur ce thème ne reflètent pas leurs compétences réelles. Une telle situation subsiste encore dans les pratiques rédactionnelles qui autorisent le choix entre une série de thèmes.
- « *La familiarité plus ou moins grande de l'élève avec le fond exerce une influence importante sur la qualité du travail* » (Hotyat, 1962, p. 54). Dans cette situation, les

apprenants qui ont des connaissances antérieures sur le thème (en contexte scolaire ou extra-scolaire) réussissent mieux leurs rédactions.

- Le milieu socio-économique dont les apprenants sont issus a des répercussions sur leurs aptitudes rédactionnelles. De ce point de vue, l'activité de rédaction privilégierait une catégorie restreinte d'apprenants issus de milieux favorisés.
- Les idées des apprenants, si elles sont différentes, pourraient biaiser la correction. Celle-ci est d'une validité incertaine vu les facteurs qui affectent la fidélité de la notation comme : la correction des copies par divers correcteurs, les effets de contraste, de fatigue...

Certaines de ces critiques semblent parfaitement justifiées. Cependant, elles ne peuvent dissimuler les avantages de cette activité rédactionnelle qui persiste dans les pratiques évaluatives des enseignants et des institutions. En effet, la rédaction ou la dissertation évaluent réellement les capacités et les compétences des apprenants. Ce sont les opérations cognitives de haut niveau (synthèse, analyse...) qui sont visées par cette activité de rédaction. Outre ce premier avantage, la dissertation assure le contrôle de la compétence de communication écrite des apprenants évalués, une compétence ciblée par tout apprentissage linguistique.

En somme, les épreuves écrites et les rédactions sont des activités qui assurent l'évaluation des connaissances et des compétences des apprenants dans des disciplines telles que les langues étrangères. Elles négligent cependant un volet considérable de l'apprentissage linguistique : l'oral. En effet, si les rédactions aboutissent à des données (notes ou remarques) déterminant, à un degré variable de rigueur, la compétence de communication écrite, elles ne pourront en aucun cas apporter des renseignements sur la compétence orale des apprenants. Cela nous amène à préférer la distinction entre examens oraux et examens écrits dans la mesure où elle est plus utile en évaluation des langues étrangères. Ces examens oraux ne sont pas omis des formes d'examens listées par Hotyat. En quoi consistent-ils ? Est-il aisé de les intégrer dans le milieu scolaire ?

### **7.1.3. Les examens oraux**

En dépit de leur large usage, les examens oraux sont d'une validité contestée pour des raisons que nous citons dans ce qui suit, en nous référant à Hotyat (1962) :

- Inégalité des chances

Les examens oraux, qui proposent des situations et sujets variés aux évalués, aboutissent souvent à des jugements d'une faible validité. En effet, certains apprenants, plus chanceux que



leurs pairs, traitent des thèmes sur lesquels ils ont des connaissances, ce qui assure l'excellence de leurs productions.

- Facteurs psychologiques et personnels

Les facteurs psychologiques (anxiété, peur...) contribuent à affaiblir la performance des apprenants dans une évaluation orale. Ce qui intensifie ces facteurs défavorables est la nature des relations qu'entretient l'enseignant avec chacun des apprenants évalués. Ainsi, les apprenants qui sont habitués à communiquer avec leur enseignant ont plus d'aisance à s'exprimer que ceux qui ne participent pas souvent dans les cours. En plus de ces facteurs qui pourraient être communs à tous, certains apprenants, sous le poids des facteurs personnels tels que la timidité, ne manifestent pas leurs compétences réelles dans ces examens qui privilégient une certaine catégorie d'apprenants : les plus audacieux.

- Difficulté d'évaluer l'oral

De nombreux auteurs tels que Dolz et Schneuwly (1998), Garcia-Debanc et Delcambre (2001)... reconnaissent la difficulté d'évaluer l'oral qui ne laisse pas de trace. Contrairement aux traces écrites des apprenants analysées à l'écrit, les productions orales des apprenants ne peuvent être réécoutées pour une correction et une notation exacte. Les exigences et les problèmes d'une évaluation exacte des productions orales sont évoqués dans la citation suivante de Garcia-Debanc et Delcambre :

Un travail sur l'oral nécessite enregistrement audio et vidéo, transcriptions et réécoutes, ce qui demande donc un matériel technique d'enregistrement et beaucoup de temps, ce qui, de plus, peut éventuellement poser des problèmes déontologiques ou juridiques avec les parents d'élèves lorsqu'on procède à des tournages dans les classes (2001, p. 7).

S'agissant des exigences, deux éléments essentiels sont repérés de la citation ci-dessus : le temps et le matériel. Même si nous admettons que les institutions consentent à accorder suffisamment de temps pour la passation et la correction des examens oraux, il est connu que le matériel que ces examens requièrent ne pourrait être aisément fourni. D'ailleurs, les enseignants, vu leur formation essentiellement didactique et pédagogique, ont souvent des difficultés à le manipuler, d'où la nécessité d'un jugement fondé sur l'impression. En plus de ces exigences, Garcia-Debanc et Delcambre (2001) abordent les problèmes qui découlent de ces enregistrements. En effet, les parents, pour des raisons multiples, pourraient s'opposer à cet enregistrement étant donné que les enseignants pourraient l'exploiter à d'autres fins.

Ces contraintes posées par les examens oraux dans le milieu scolaire que nous venons de citer ne constituent aucunement une liste exhaustive. Cependant, elles contribuent à

comprendre la difficulté d'organiser des examens oraux, d'en assurer le bon déroulement et d'aboutir à des résultats valides.

## 7.2. Les tests

Selon Riba et Mègre, qui opposent test et examen, le test « *ne conduit pas à un diplôme mais à une attestation, valable pour une durée limitée dans le temps* » (2014, p. 91). Cette opposition insiste essentiellement sur le résultat auquel aboutit l'évaluation qui aura des fins différentes. Ainsi, selon ces auteurs, contrairement aux examens qui sanctionnent un cycle d'apprentissage, les tests aboutissent à un descriptif des connaissances et de compétences des évalués à un moment donné (le moment du test). Ce moment du test a permis à Mothe, dans son ouvrage « *L'évaluation par les tests dans la classe de français* », d'introduire une typologie de tests : les tests de pronostic et les tests de diagnostic.

Les premiers, introduits à des fins prédictives et diagnostiques, sont répartis en deux catégories : les tests d'aptitude qui « *doivent fournir une indication statistique des probabilités de succès dans un type déterminé d'apprentissage* » (Mothe, 1975, p. 17) et les tests de niveau. Ceux-ci ne visent pas la prévision de la réussite des évalués ; ils aboutissent à des résultats qui servent à sélectionner les apprenants ou à les répartir sur des classes selon leurs besoins. Dans ce cas-ci, des contenus précis seront dispensés à chaque classe pour le développement de leurs compétences. Alors, les tests de pronostic fournissent des indications sur les apprenants, leurs compétences et leurs besoins pour une intervention appropriée. Leur fonction est celle de l'évaluation diagnostique introduite en amont des apprentissages.

Les seconds tests, ou tests de diagnostic, se situent en aval des apprentissages. Contrairement aux précédents, ces tests contrôlent et vérifient les apprentissages des apprenants suite aux enseignements dispensés. Ils sont divisés en deux catégories : les tests de contrôle et les tests de progrès. La première catégorie de tests vise le contrôle des acquis au bout d'une série considérable d'apprentissages, d'où leur conception à grande échelle. Quant à la deuxième catégorie, elle regroupe les tests conçus par l'enseignant lui-même à des fins de régulation. Ce qui a amené Mothe à leur attribuer une fonction essentiellement formative, tandis que la fonction sommative est assignée aux tests de contrôle.

Ce ne sont pas ces deux fonctions qui interpellent notre attention à ce niveau vu l'effondrement de frontières entre les diverses formes d'évaluation. En effet, une évaluation sommative, introduite au terme d'un cycle de formation, est elle-même diagnostique -dans la mesure où elle détermine les apprentissages ultérieurs à programmer- et formative car elle contribue à la régulation des acquis. Ce sont notamment les liens qui pourraient exister entre

certaines sous-types et examens. Ainsi, les tests de contrôle introduits au bout des apprentissages d'une année ou d'un cycle se rapprochent des examens scolaires, évoqués supra. Cependant, la standardisation requise des premiers à tous les niveaux de conception, de passation et de correction les distinguent des seconds.

En considérant tout ce qui précède, nous trouvons que la distinction établie entre les examens et les tests s'appuie sur deux éléments essentiels : le moment de leur organisation et les résultats auxquels ils aboutissent (diplôme ou attestation). Alors que ces critères distinctifs existent, la confusion entre les deux termes est toujours perceptible vu la nature des épreuves que les examens et les tests proposent : les mêmes épreuves pourraient être retenues dans les deux instruments (tests et examens). A ce point commun entre ces deux instruments, s'ajoute la similitude de leurs sous-types. Ainsi, les examens de fin d'études évoqués par Hotyat correspondent aux tests de contrôle abordés par Mothe. Aussi, les deux auteurs associent la fonction pronostique à un des sous-types de leurs classifications.

### **7.3. Pour une évaluation scolaire : tests ou examens**

Les précisions fournies sur les tests et les examens, deux instruments d'évaluation, permettent certainement de comprendre leurs particularités. Cependant, notre objectif ne se limite pas à la compréhension de leurs caractéristiques. En effet, à travers la distinction des deux modes, nous tenterons de répondre à une question cruciale dans notre étude : lesquels des deux instruments évoqués supra une évaluation scolaire introduit-elle ? Les connaissances et les compétences des apprenants sont-elles évaluées par des tests ou par des examens ?

Tests et examens sont des instruments évaluatifs qui aboutissent à une certification, qu'elle soit de courte ou de longue durée. Si les premiers préservent leur caractère certificatif, les seconds l'ont délaissé. Ainsi, les examens ne sont plus associés à une certification précise. Leur introduction est fréquente dans les univers scolaires pour dresser le bilan des acquis réellement maîtrisés par les apprenants-évalués, au bout d'une série d'apprentissages. Ces examens scolaires, qui révèlent les compétences réelles des apprenants évalués, sont intégrés pour évaluer les diverses disciplines et matières scolaires. La réussite de ces apprenants est conditionnée par les résultats obtenus dans toutes ces matières.

Distinct du domaine scolaire, le domaine de la formation recourt notamment aux tests : des tests de contrôle qui sont couronnés par des attestations remises aux évalués et des tests de pronostic qui ont une fonction sélective. Ces derniers tests se sont multipliés récemment. Ainsi, nous assistons à une prolifération des tests de langues étrangères qui ouvrent la voie aux formés pour poursuivre leurs études dans d'autres pays et pour décrocher des postes de travail (insertion

professionnelle). Dans ce cadre, de nombreux auteurs tels que Gruca et Cuq (2008), Noel-Jothy et Sampsonis (2006) énumèrent dans leurs écrits les tests de Français Langue Etrangère (DELFL, DALF...). Ces tests, qui ont certainement inspiré les examinateurs dans les univers scolaires et amélioré leurs pratiques, ne sont introduits que dans les formations complémentaires ou professionnelles. Mentionnons que le concept de test, plus récurrent dans le domaine de la formation, est également employé pour évoquer une évaluation scolaire. Ainsi, Mothe dans son ouvrage « *L'évaluation par les tests dans la classe de français* », cite les tests de classe qui assurent le contrôle du degré d'assimilation des cours.

Réduire les évaluations sommatives dans le milieu scolaire aux examens est certainement exagéré. En prenant en considération les idées de certains auteurs comme Pelpel (1986), Riba et Mègre (2014)..., nous trouvons que le terme *examen* ne correspond qu'aux bilans de fin de cycle. Les autres évaluations sommatives qui sont introduites en aval d'un cours, d'un trimestre, d'un semestre ou d'une année prennent des appellations diverses dans les systèmes éducatifs. Dans le système éducatif algérien, les *devoirs* closent les séquences d'apprentissages tandis que les *compositions* correspondent à la fin des trimestres qui constituent l'année scolaire. Devoirs ou compositions, ces évaluations sont des *épreuves*, un terme qui nous semble plus approprié pour désigner les évaluations sommatives dans le milieu scolaire, notamment que les caractéristiques particulières aux tests et aux examens en limitent les contextes d'emploi.

Tenant compte de ce qui précède, nous pouvons associer les tests à l'évaluation certificative et lier les examens à l'évaluation sommative, bien que ces examens ne soient liés -par des auteurs comme Pelpel- qu'aux évaluations de fin des cycles d'études. En effet, notre étude considère l'évaluation certificative comme une évaluation sommative qui aboutit, en plus de la notation et du jugement, à une certification. Nous précisons que les termes employés pour désigner les modalités adoptées en évaluation sommative sont d'une variété inouïe. Examens ou tests, ces évaluations sont des *épreuves* qui varient des exercices structuraux classiques aux situations d'intégration complexes. Quelles sont donc les épreuves qui pourraient être introduites dans les examens de langues étrangères, plus particulièrement du FLE ? La réponse à cette question, nous la reporterons au premier chapitre de notre partie pratique dans lequel les évaluations sommatives du FLE du cycle moyen sont analysées, en vue de déterminer les activités auxquelles les enseignants recourent.

### **Conclusion**

Notre premier chapitre théorique qui porte sur l'évaluation, un concept-clé de notre étude, retrace l'évolution de la signification de ce concept à travers les années. Désignant un processus confondu avec la mesure et complètement dissocié de l'apprentissage scolaire dans le paradigme ancien, le terme évaluation renvoie dans le paradigme nouveau à une pratique distincte de la mesure, qui en constitue une composante essentielle, et étroitement liée aux apprentissages scolaires. L'éclaircissement de ce changement sémantique fournit certes des précisions précieuses sur le concept d'évaluation. Cependant, il ne se substitue guère à sa définition, que nous avons formulée dans notre étude en nous référant à la littérature relative sur le thème. Les définitions citées privilégient des aspects particuliers de ce processus : certaines le considèrent comme un jugement de valeur, d'autres insistent sur la collecte des données dans le processus évaluatif, quelques-unes l'associent à la notation... Quel que soit le point de vue privilégié, toutes ces définitions sont extrêmement utiles car elles explicitent les composantes qui constituent l'évaluation ainsi que les outils que cette opération investit. D'ailleurs, une définition satisfaisante requiert l'intégration de tous les points de vue abordés.

Bien que les éléments cités soient suffisants pour comprendre l'évaluation ainsi que le rapport que celle-ci entretient avec l'apprentissage, nous avons jugé nécessaire la précision des acteurs évaluateurs et évalués pour déterminer les contours de l'évaluation dont il est question dans la présente étude, et ce, en raison de la diversité des évaluateurs (inspecteurs, enseignants, apprenants...) et celle des évalués (apprenants, enseignants, programmes...). Ayant précisé que notre étude porte sur l'évaluation des apprenants, plus précisément de leurs compétences en FLE, par les enseignants, une nouvelle précision était indispensable : la forme évaluative privilégiée par l'étude. Dans ce contexte, nous avons précisé que de nombreux types d'évaluation sont proposés et que ceux définis sont les plus courants (évaluation diagnostique, formative et sommative). Soulignons que ces trois types sont intimement liés aux fonctions de l'évaluation.

L'évaluation sommative que nous avons choisie comme thème de notre recherche a suscité de nombreuses questions auxquelles nous avons tenté de répondre. Sa définition, premier élément traité, tient compte de plusieurs critères : le moment de son introduction, ses buts, son résultat, ses approches et ses outils. En effet, l'évaluation sommative est introduite au terme des apprentissages pour contrôler les acquis des apprenants et établir un bilan des apprentissages ; elle peut être critériée (basée sur des critères et indicateurs précis) ou normative (fondée sur le classement des apprenants) ; elle aboutit à un jugement définitif et à la notation qui ne traduit en aucun cas les compétences ou les connaissances des évalués. Certains de ces

critères ont permis la distinction de cette forme de la forme diagnostique et sa confusion avec la forme certificative. Si la distinction entre les premières formes citées est admise en dépit de la fonction régulatrice que la forme sommative pourrait remplir, nous rejetons la confusion avec la forme certificative qui, contrairement à la forme sommative, aboutit à des certifications inexistantes dans le milieu scolaire. Ceci illustre la différence des fonctions associées à ces deux formes. Mentionnons que tous les éléments abordés sur l'évaluation ne peuvent répondre à toutes les questions imposées par la complexité du thème de l'évaluation. Dans ce sens, De Ketele affirme : « *Plus on pénètre dans le domaine de l'évaluation, plus on prend conscience du caractère encyclopédique de notre ignorance, plus nos certitudes sont remises en question. Toute question posée entraîne d'autres. Chaque arbre en cache un autre et la forêt paraît immense* » (De Ketele, Préface, cité dans Cardinet, 1986, p. 6.).

En plus de ces éléments qui définissent la forme sommative d'évaluation, nous avons abordé les critères qui assurent la professionnalisation de l'évaluation sommative. Celle-ci, se concrétisant dans le milieu scolaire à travers des examens, des tests, des épreuves ou des rédactions, exige la validité, la fidélité, l'objectivité, la faisabilité, l'équité... Ces critères non-exhaustifs assurent des résultats reflétant le plus fidèlement possible les compétences des apprenants évalués. Tous ces critères sont définis dans ce chapitre selon les points de vue de plusieurs auteurs : la fidélité renvoie à la précision de la mesure qui se manifeste par la stabilité des résultats, l'objectivité désigne la stabilité des résultats en dépit de la diversité des correcteurs... A travers la définition de ces critères, nous avons constaté que certains sont plus importants que d'autres. En effet, la validité et la fidélité sont des critères majeurs dont les autres en constituent des dimensions. Néanmoins, de ces deux critères, la validité est le critère le plus important dans la mesure où elle englobe les autres, notamment que la fidélité est rangée parmi les types de validité. Réalisant l'importance de ce critère qui conditionne la réussite du processus évaluatif, notre recherche est centrée sur la validité de l'évaluation sommative dans le système éducatif algérien.

Constituant un concept clé de notre recherche, le concept de validité, défini brièvement dans ce chapitre, sera traité dans le chapitre suivant en vue de donner des précisions plus utiles à notre étude. En effet, la définition de ce critère permettra de repérer quelques indicateurs qui serviront ultérieurement à valider les outils évaluatifs ou de jugement ; et même de concevoir des outils valides notamment avec la réforme de la pédagogie en Algérie.

# **CHAPITRE 2 : LA VALIDITE ET LA VALIDATION DES EPREUVES**

## Introduction

L'évaluation est un processus qui occupe une importance majeure dans l'enseignement/apprentissage de toutes les disciplines. A travers ses multiples fonctions, ce processus soutient l'apprentissage scolaire dans tous ses moments. Introduite en amont, l'évaluation diagnostique assure la planification de contenus linguistiques et culturels adaptés aux besoins des publics enseignés. Étant donné que les programmes et les contenus scolaires réservés à chaque niveau scolaire sont institutionnellement préétablis, cette forme évaluative se voit attribuée une nouvelle fonction : celle d'adapter les contenus déjà retenus afin qu'ils soient plus conformes à la particularité des apprenants. Lorsqu'elle est intégrée pendant les apprentissages, l'évaluation assure une bonne assimilation des contenus enseignés : elle remédie aux lacunes constatées en vue de garantir le transfert de ce qui est appris dans de nouveaux contextes authentiques. Cette deuxième fonction assurée par l'évaluation est remplie par la forme formative qui revêt actuellement une importance indéniable. En dépit de la valeur de cette forme, la forme sommative -dont les retombées sont considérables sur les plans personnel, professionnel et social- a essentiellement retenu notre attention pour constituer le sujet de notre recherche.

L'évaluation sommative s'étale sur plusieurs étapes qui aboutissent à un jugement. Ce jugement, contrairement à ce que son nom l'indique, doit être exempt de subjectivité : il s'agit du jugement professionnel qui s'appuie sur la mesure, une étape aussi décisive que le jugement. En effet, ce sont les données auxquelles les outils de la mesure aboutissent qui assurent la validité du jugement, fondé également sur des outils pertinents. Mentionnons que des critères professionnels sont requis pour assurer la validité de la mesure et celle du jugement. Il s'agit de : la validité, la fiabilité, l'équité, l'objectivité... Tous ces critères ont été brièvement expliqués dans notre premier chapitre théorique. Néanmoins, à travers les multiples définitions citées, il s'est avéré que la validité constitue la qualité la plus importante (Angoff, 1988) et la plus difficile à assurer en évaluation, comme Mothe (1975) le précise, d'où l'intérêt que notre présent travail lui accorde.

L'importance de ce critère de validité est indéniable : il s'agit d'un critère général englobant les autres critères auxquels il est souvent opposé (la fidélité, l'objectivité...). Ce lien que la validité entretient avec les autres critères sera largement expliqué dans le présent chapitre pour préciser quelle conception nous adopterons dans le reste de notre étude. En effet, il est nécessaire de mentionner que si certains rangent tous les critères au sein du critère global de la validité, d'autres continuent à établir une distinction entre les divers critères attribuant à chacun



un aspect particulier. Aborder cette opposition contribue considérablement à mieux cerner la notion de validité, qui constitue un concept-clef dans notre étude, et à écarter la confusion terminologique qui l'entoure. Cette confusion apparaît dans la diversité des conceptions définitoires adoptées par les auteurs. Ainsi, les premiers (Tagliante, 2005) continuent à percevoir la validité comme une capacité des outils évaluatifs à mesurer réellement ce qui est évalué, tandis que les seconds (Kane, 2006) l'associent plutôt à la qualité de l'interprétation des résultats obtenus. Ces deux tendances seront détaillées dans les éléments qui suivent pour arriver à une conception précise de la validité. Cette précision est indispensable pour déterminer ultérieurement les critères et les indicateurs assurant la validation des épreuves évaluatives retenues.

En dépit de cette importance de la validité, celle-ci demeure difficile à assurer dans les outils évaluatifs à intégrer, notamment qu'elle tisse un lien étroit avec la pédagogie appliquée, les objectifs à évaluer, la représentativité des contenus... C'est en tenant compte des exigences de la pédagogie appliquée que les outils évaluatifs sont conçus puis intégrés. La conception des dispositifs évaluatifs constitue un axe considérable de notre recherche, mais nous n'en aborderons pas toutes les étapes : seule la validation sera abordée à ce niveau vu le lien que celle-ci entretient avec la validité.

Les deux notions de validité et de validation sont distinctes étant donné que la seconde renvoie au processus d'expérimentation aboutissant au jugement de validité. Ce sont ces procédures qui contribuent à la validation des épreuves conçues pour évaluer la maîtrise des objectifs visés par les programmes d'études que nous exposerons dans ce chapitre. Cet aperçu nous permettra de déterminer celles que nous pourrions investir pour évaluer la validité des dispositifs intégrés dans le système éducatif pour évaluer les compétences des apprenants algériens en FLE. Nous aurions pu traiter cette validation parmi les étapes de la conception des dispositifs évaluatifs en éducation, mais son autonomie qui assure son application à des dispositifs existants nous a décidée à la traiter profondément dans ce chapitre.

Nous tenterons alors de répondre à plusieurs questions en nous appuyant sur les références que nous avons pu collecter sur la validité et sur les procédures de validation. Parmi ces questions, nous citons : en quoi consiste la validité ? Quels rapports entretient-elle avec les autres critères ? Quelles sont les procédures de validation qui existent ? Lesquelles seront utiles au contexte de notre étude ?

## 1. La validité : des conceptions définitoires

La notion de validité en évaluation n'est pas récente. Newton (2012) - qui en a retracé l'évolution – cite les premiers chercheurs tels que Terman (1915), Starch (1916), Thorndike (1916) et Hartog (1918) qui ont abordé cette notion dont les définitions sont diverses<sup>28</sup>.

La diversité des définitions, que nous citerons ci-dessous, formulées par des auteurs tels que C. Tagliante (2005), P. Riba et B. Mègre (2015), Noël-Jothy et B. Sampsonis (2006)... nous conduisent à affirmer la difficulté de proposer une définition unique et complète qui engloberait toutes les acceptions évoquées par les chercheurs. En effet, le concept de validité est présenté tantôt comme une qualité liée au test, tantôt comme une caractéristique du score ou du résultat obtenu suite à la passation de celui-ci. Chez certains auteurs, il est opposé aux autres critères comme la fiabilité professionnelle alors que chez d'autres il est conditionné par ces mêmes critères.

Sans nous étaler sur ces dichotomies qui seront abordées dans le reste de notre travail, nous mentionnons qu'à la suite des travaux de plusieurs auteurs sur la validité (Buckingham, Kelley, Angoff, Guilford, Gronbach...), nous retenons deux tendances qui se présentent comme contradictoires : la première est présentée comme ancienne et restreinte alors que la seconde est perçue comme plus récente et plus large. Dans ce qui suit, nous présentons chacune des deux catégories. Mentionnons que la conception que nous avons qualifiée de traditionnelle et ancienne ne cesse plus d'être investie par les chercheurs.

### 1.1 La validité : une qualité du test

Nous devons la première définition de la notion de validité à Buckingham et al. Selon eux, la notion se rapporte aux problèmes « *qui ont trait à la détermination de ce que le test mesure* » (1921, 80, Cités dans André, Loye & Laurencelle, 2015, p. 131). Cette définition, qui est adoptée par d'autres auteurs tels que Kelley (1927), annonce une conception classique<sup>29</sup> de validité liée au test qui persiste même dans les définitions récentes.

Dans « *Certifications et Outils d'Évaluation en FLE* », les auteurs Françoise Noël-Jothy et Béatrix Sampsonis définissent la validité comme « *la capacité d'un examen à refléter les*

---

<sup>28</sup> L'évolution de la notion de validité, telle que Newton la retrace, est abordée par N. André, N. Loye et L. Laurencelle (2015) dans leur article « La validité psychométrique : un regard global sur le concept centenaire, sa genèse, ses avatars », publié dans la revue *Mesure et évaluation en éducation*.

<sup>29</sup> Dans les travaux classiques, les auteurs associent la validité à ses types particuliers. Ainsi, une partie importante de « *Educational measurement* » aborde la validité du critère. Cronbach (1971) met en valeur la validité du contenu et celle du concept (Raïche, Paquette-Côté & Magis, 2011).

*compétences de l'évalué*» (2006, p. 23). Ainsi, les résultats obtenus sont interprétés de façon à révéler non seulement les connaissances, mais également les compétences effectives des apprenants à travers la généralisation du score obtenu et la sélection d'items représentatifs. Si la validité dans cette définition est associée au test, elle est liée chez Christine Tagliante à l'instrument de mesure ou à l'activité évaluative. En effet, l'auteure avance que la validité est « *la qualité d'une activité d'évaluation qui fait que cette activité mesure exclusivement et exactement ce qu'elle est censée mesurer ce pour quoi elle a été conçue* » (2005, p. 23). Liées à la mesure, les notions d'exclusivité et d'exactitude<sup>30</sup> évoquées par Tagliante ne s'obtiennent que par la congruence entre l'instrument évaluatif et l'objectif déjà ciblé par l'enseignement/apprentissage. De même, cet instrument doit permettre la collecte de données informatives et utiles à l'interprétation exacte des résultats.

Patrick Riba et Bruno Mègre rejoignent Tagliante dans sa conception de la validité en évaluation en l'associant à l'activité évaluative ou l'item sélectionné dans un test. Ils avancent que « *la validité doit être respectée en contexte de conception de tâche, d'exercice ou d'activité d'évaluation afin de garantir le degré de mesure de ce que l'on est censé mesurer* » (Riba & Mègre, 2015, p. 63). Outre le lien entre item et objectif sur lequel les auteurs insistent, ceux-ci montrent la nécessité du respect de ce critère lors de la conception de dispositifs évaluatifs pour une mesure exacte des objectifs éducatifs visés. Néanmoins, il ne s'agit pas de l'unique définition que les auteurs proposent. En effet, dans le glossaire du même ouvrage « *Démarche Qualité et Evaluation en Langues* », nous retrouvons une deuxième définition du concept validité. Cette fois, celui-ci n'est plus lié à l'instrument de mesure mais plutôt à la note, résultat obtenu suite à la passation du test. Ainsi, ils présentent la validité comme le « *degré auquel les notes d'un test permettent de tirer des conclusions appropriées, significatives et utiles en relation avec l'objet du test* » (Riba & Mègre, 2015, p. 212). La validité se rapporte ainsi à la représentativité et à l'utilité des résultats d'une évaluation donnée qui sont interprétés de façon à refléter le degré de maîtrise des contenus retenus. Parmi toutes les définitions que nous avons évoquées ci-dessus, celle-ci se rapproche davantage de la tendance récente de la validité qui est liée à l'interprétation du score obtenu.

---

<sup>30</sup>Pour Tagliante, trois notions de base sont liées à la validité à savoir : mesure, exclusivité et exactitude. L'auteure fournit une explication de la deuxième notion dans son ouvrage *L'évaluation et le cadre européen commun* (2005). Selon elle, l'exclusivité est résultante de « *la correspondance entre le contenu de l'activité d'évaluation et l'objectif opérationnel visé* » (2005, p. 23).

Les définitions que nous avons classées dans la conception traditionnelle de la validité ne se réduisent pas à une caractéristique du test ou de l'instrument évaluatifs. Elles se rapportent également à certains de ses indicateurs tels que la « *cohérence entre l'épreuve et l'objectif* » (Huver & Springer, 2011, p. 25). Huver et Springer, en plus du lien qu'ils établissent entre validité et épreuve, intègrent une notion incontournable en évaluation scolaire, celle des objectifs. Loin du débat sur la limite du concept dans une évaluation de compétences et de sa particularité à la pédagogie par objectifs, son intégration a permis d'établir un indicateur de validité et de validation des activités d'évaluation. En effet, la détermination des objectifs contribue à la conception des cours, mais aussi à l'élaboration de tests (Morissette, 1993) dans la mesure où nous évaluons ce qui a été enseigné aux apprenants et ce qui se concrétise dans des objectifs précis. Soulignons que les objectifs sélectionnés doivent être représentatifs du programme éducatif retenu.

La représentativité en évaluation est une notion-clé que nous retrouvons dans d'autres ouvrages. Ainsi, dans « *Réforme de l'Education et Innovation Pédagogique en Algérie* », la validité est définie comme « *la représentativité de l'échantillon d'items par rapport à l'univers de référence* » (Gérard, 2006, p. 90). Ainsi, les items proposés dans un examen ne constituent qu'un échantillon de l'ensemble des cours programmés que nous désignons par *univers de référence*. La représentativité de cet échantillon conditionne la validité de l'évaluation passée : à travers celui-ci, les enseignants-évaluateurs devraient avoir un aperçu sur le degré de maîtrise du programme éducatif. La représentativité des items est étroitement liée à celle des objectifs étant donné que l'élaboration des premiers s'appuie sur la formulation et la précision des seconds. Ce rapport étroit qu'entretiennent item et objectif est investi dans le même ouvrage où la validité est encore définie en tant que « *degré d'adéquation entre ce que l'on déclare faire (évaluer telle ou telle dimension) et ce qu'on l'on fait réellement* » (Gérard, 2006, p. 96). La congruence entre les objectifs de l'évaluation et son contenu dont il est question dans ce contexte n'est autre que celle entre les items retenus et l'univers de référence. En effet, les objectifs se traduisent dans des items ou des activités évaluatives alors que le contenu est un échantillon représentatif du programme.

En plus des définitions citées ci-dessus, certains auteurs tels que Gilbert De Landsheere (1992) définissent la validité par ses différents types (validité du contenu et validité prédictive). D'autres chercheurs (Hotyat, 1962) insistent dans leur tentative de définir le concept sur son rapport avec les autres critères professionnels d'évaluation. A cet égard, Hotyat affirme que « *la fidélité de l'examen est une condition nécessaire à sa validité* » (1962, p. 30). Il en est de

même pour Cureton (1951) qui définit la validité par les deux dimensions qui lui sont assignées : la pertinence et la fiabilité. Cet auteur associe le critère de validité à celui de fiabilité ou fidélité qui ne constitue qu'une dimension du premier en dépit de la distinction établie entre les deux critères.

En somme, ces définitions que nous avons énumérées en dépit de quelques différences renvoient toutes à la validité en tant qu'une qualité du test. Mais, sa définition a évolué car « *on la considère aujourd'hui non plus comme une caractéristique intrinsèque de la mesure* » (Raïche, Ndinga & Meunier, 2013, p. 1) : elle est plutôt liée à l'interprétation des résultats. A la différence de l'acceptation traditionnelle de la validité, exposée supra et qualifiée de *fausse* par des auteurs comme Raïche, Ndinga et Meunier (2013), une seconde acception plus récente est explicitée dans ce qui suit.

### **1.2. La validité : une caractéristique de l'interprétation des résultats**

La conception classique de la validité en tant que qualité du test commence à s'élargir donnant ainsi naissance à une conception plus récente. Selon André, Loye et Laurencelle, « *Gronbach (1971) a notamment cristallisé cet élargissement de perspective en affirmant que l'on ne valide pas le test, mais l'interprétation des mesures qu'il fournit* » (2015, p. 132). Ainsi, cette notion ne devient pas seulement une propriété de test, mais elle est également liée à l'interprétation des résultats. Comme plusieurs interprétations sont possibles en évaluation, il devient nécessaire de valider une interprétation jugée acceptable par les évaluateurs.

Dans la même perspective, Raïche, Ndinga et Meunier, qui réfutent la conception classique de la validité, l'associent plutôt à l'interprétation des résultats. Selon eux, elle est « *en relation avec l'utilisation et l'interprétation du score associé à la mesure* » (2013, p. 1). Ainsi conçue, la validité n'est plus liée à la capacité du test qui doit mesurer efficacement ce qu'il est destiné à mesurer : elle est notamment une caractéristique de l'interprétation des scores. C'est pour cette raison que les recherches récentes préfèrent le concept validité de l'interprétation à celui de validité de test.

Cette évolution de significations dont il est question résulte en réalité d'un changement de regard envers les divers moments de l'évaluation. Si, dans la conception classique, la phase de la mesure détient l'intérêt des auteurs, cette importance s'est amoindrie dans la conception nouvelle qui rehausse la phase du jugement au premier plan. Ce changement de position est perceptible dans certaines définitions de la validité qui l'associent au jugement. Tel est le cas de Messick (1989)-cité par Karine Paquette-Côté et Gilles Raïche- qui considère la validité

comme « *le jugement évaluatif intégré du degré auquel la preuve empirique et le rationnel théorique supportent la justesse et la pertinence des inférences et des actions basées sur les résultats d'une évaluation* » (1989, p. 13, cités dans Paquette-Côté & Raïche, 2013, p. 8), une définition récemment réfutée par Borsboom, Mellenbergh et Van Heerden (2004). Pour ces auteurs, cette définition ne s'écarte pas de celle de Cronbach (1980) qui associe la validité à un argument. Cet argument, qui s'appuie sur des preuves empiriques (étude de la manifestation concrète de la compétence des apprenants) et sur des données théoriques (des modèles scientifiques), est constitué de plusieurs inférences<sup>31</sup>. Cette notion d'argument, nous la retrouvons dans les travaux de Kane (1990, 2006), qui associe la validité de l'interprétation au degré de plausibilité et de l'argument interprétatif<sup>32</sup>. Selon lui, cet argument est constitué des inférences et des suppositions impliquées par l'interprétation pour passer des scores du test à des conclusions ou des décisions appropriées (Kane, 1990)<sup>33</sup>. La pertinence, la cohérence et la plausibilité de ce premier argument sont évaluées à travers un second type d'arguments, l'argument de validité, pour la validation la validation des interprétations (Cubilo, 2014).

Cette conception de la validité de l'évaluation en tant qu'argument interprétatif est à l'origine de plusieurs modèles de validation dont les deux principaux, selon Paquette-Côté et Raïche (2011), sont :

- Evidence-Based Design<sup>34</sup> : conçu par Mislevy et ses collaborateurs (Mislevy, Almond & Lukas, 2004 ; Mislevy & Haertel, 2006 ; Mislevy, Steinberg, Almond & Lukas, 2006), ce modèle est basé sur la preuve empirique,
- Argumentation interprétative : développé par Kane (2006), ce second modèle est axé sur le rationnel théorique.

La définition de la validité, que Messick (1989) propose, n'est pas la seule qui regroupe l'empirique et le théorique. En effet, nous retrouvons dans *Manuel pour l'élaboration et la passation de tests et d'examens de langue* une définition de la validité en tant que « *degré de preuves et de théorie sous-tendant l'interprétation des scores entraînée par les utilisations*

<sup>31</sup> Les inférences sont des raisonnements qui s'enchaînent pour aboutir à des conclusions appropriées.

<sup>32</sup> Dans son document intitulé «An Argument-based Approach to Validation», Kane avance : « *The validity of an interpretation can be defined in terms of the degree to which the interpretive argument is plausible and appropriate* » (1990, p. 9)

<sup>33</sup> Pour définir l'argument interprétatif, nous nous référons aux deux citations suivantes de Kane : « *The inferences and assumptions constitute an argument* » (1990, p. 2), et « *The interpretation of test scores involves inferences from the test scores to various conclusions, and possibly, to decisions about appropriate actions* » (1990, p. 1).

<sup>34</sup>Validation centrée sur les preuves en français.

*données des tests*» (AERA, APA, NCME, 1999<sup>35</sup>, Cité par Conseil de l'Europe, 2001, p. 16). Autrement dit, l'interprétation des scores ou résultats obtenus devrait d'appuyer sur des preuves empiriques et théoriques aboutissant à un jugement valide sur les capacités et les compétences linguistiques des apprenants ainsi que sur l'utilisation de la langue dans diverses situations.

Bien qu'aucune évaluation ne puisse être authentique au point de déterminer les compétences réelles des évalués, les résultats du test, qui représentent la compétence observée de l'apprenant, sont interprétés de façon à prédire sa compétence dans la vie réelle étant donné que la performance de l'apprenant dans les tâches évaluatives est considérée comme un argument de sa compétence dans le monde authentique. Dans ce sens, la validité « *fait un lien entre la performance dans la réalisation de la tâche et la preuve de la compétence langagière du candidat dans le monde réel* » (AERA, APA, NCME, 1999, Cité par Conseil de l'Europe, 2001, p. 16), d'où la nécessité de veiller à une élaboration étudiée des tests pour des interprétations significatives. De même, la signifiante et l'utilité des interprétations auxquelles un test aboutit en attestent la validité. Mais, la relativité des interprétations complique tout jugement de validité de test. Dans ce contexte, Gronbach précise « *que puisque chaque degré d'interprétation a son propre degré de validité on ne peut jamais arriver à la simple conclusion qu'un test particulier est valide* » (Cronbach, 1971, p. 447, Cité dans Paquette-Côté & Raïche, 2013, p. 8).

Cette conception de la validité, en dépit de sa complexité, semble plus précise et plus fructueuse que celle précédemment évoquée (la validité en tant que qualité du test). Définie comme le reflet de la compétence des apprenants, celle-ci demeure limitée dans la mesure où aucun test ou examen ne pourra donner un aperçu fidèle sur les compétences des apprenants. Ce n'est qu'avec l'interprétation des résultats obtenus que lesdites compétences sont décrites. Cependant, Borsboom, Mellenbergh et Van Heerden (2004), qui évoquent les deux perspectives définitoires, réfutent la définition de la validité associée à une interprétation des scores pour revenir à la définition de Kelley (1927) que nous avons déjà citée. Ils affirment que « *la validité est un concept très basique et était correctement formulé, par exemple, par Kelley (1927, p. 14) lorsqu'il a déclaré qu'un test est valide s'il mesure ce qu'il prétend mesurer* »<sup>36</sup>. (2004, p. 1061).

---

<sup>35</sup> La page n'est pas citée dans la source de la citation.

<sup>36</sup>Traduction de « *It is a very basic concept and was correctly formulated, for instance, by Kelley (1927, p. 14) when he stated that a test is valid if it measures what it purports to measure* ».

En somme, nous devons mentionner que la validité de l'interprétation correspond à une conception plus profonde et plus utile de la validité étant donné qu'elle est à l'origine de plusieurs modèles de validation comme ceux de Mislevy et de Kane que nous développerons dans ce qui suit. Néanmoins, la perspective jugée fautive et dépassée nous semble indispensable à prendre en considération. En effet, les résultats du test ne peuvent être valides si celui-ci ne l'est pas. Ceci nous conduit à prendre la validité dans une perspective plus large englobant tous les moments du processus évaluatifs : la conception du test, son administration, l'interprétation de ses résultats, les décisions... De tous ces moments néanmoins, seules les phases de mesure et de jugement semblent privilégiées par la recherche scientifique, car elles sont directement liées aux fonctions des enseignants qui doivent élaborer des outils assurant la validité de ces étapes. Dans ce contexte, Scallon énumère deux composantes essentielles du processus évaluatif : « *les situations d'évaluation (tâches) et les outils de jugement* » (2004, p. 261), deux composantes sur lesquelles notre étude insiste dans la mesure où nous concevons la validité comme la capacité de tous les outils investis tout au long du processus évaluatif à fournir un aperçu assez fidèle des compétences des apprenants évalués à un moment précis (étant donné que les compétences sont évolutives).

## 2. Les formes (types) de la validité

En plus des diverses définitions que les chercheurs ont établies de la notion de validité, ceux-ci en ont proposé des types variés. Ainsi, André, Loye et Laurencelle (2015) récapitulent-ils ces formes (ou types) multiples dans un tableau qui englobent les plus connues comme : la fidélité, la validité manifeste, concomitante, prédictive, de contenu, de construit, unifiée... Pour sa part, De Ketele mentionne cette infinité de formes dans la citation suivante :

En voici une liste non exhaustive : validité de construit, de construct, hypothético-déductive, structurelle, de contenu, de différenciation, curriculaire, discriminante, logique, définitoire, empirique, corrélationnelle, critérielle, factorielle, concourante, concomitante, apparente, de statut, congruente, convergente, prédictive, écologique, statistique... (2010, p. 29).

Dans notre travail, nous n'aborderons que quelques types établis par certains auteurs. Ainsi, Landsheere (2000) distingue la validité du contenu de la validité prédictive. Pour sa part, Tardif (1993) évoque la validité du contenu, la validité écologique et la validité déontologique. La typologie établie par Demeuse et Henry (2004) comporte cinq formes : la validité prédictive, la validité du contenu, la validité du construit, la validité concourante et la validité



incrémentielle. Nous retrouvons d'autres types tels que la validité du critère, la validité concomitante...

Nous tenterons dans ce qui suit d'expliquer les différents types établis par les chercheurs précédemment cités en prenant en considération les points de vue de divers chercheurs. Il importe de préciser que cette typologie présente à nos yeux un double intérêt. D'une part, elle permet de mieux cerner le concept de validité et d'en comprendre les diverses dimensions. D'autre part, elle contribue à mieux expliciter les procédures de validation inspirées ou fondées sur ces types.

### **2.1. La validité du contenu**

Issue de la validité manifeste qui existe depuis 1940, la validité du contenu s'impose en 1950 (André, Loyer & Laurencelle, 2015). Elle est désignée par plusieurs concepts : apparente, manifeste, échantillonnale...

Afin de définir ce premier type, Tardif (1993) reprend Legendre (1985). Selon celui-ci, il s'agit du « *type de validité qui indique le degré de : a) congruence de chacun des items avec l'objectif mesuré ; b) représentativité de l'ensemble des items d'un instrument de mesure en regard de ce qu'on veut mesurer* » (Legendre, 1985, p. 634, Cité dans Tardif, 1993, p. 07). Cette première définition insiste sur la corrélation entre les items et les objectifs d'évaluation et sur leur représentativité de l'univers de généralisation (le programme d'études). En effet, les items retenus dans une épreuve ou un test doivent nécessairement constituer un échantillon représentatif du programme pour que le jugement de compétence exprimé à partir de ces items puisse être à raison généralisé.

Relative au contenu du test, cette validité est également désignée par validité manifeste ou apparente (Demeuse & Henry, 2004). Par validité manifeste, Laurencelle (1998) désigne la correspondance du contenu du test avec ce que nous voulons mesurer, une conception qui rejoint celle de Legendre.

Selon De Landsheere, cette validité « *dépend de la qualité et de l'adéquation de l'échantillonnage des connaissances et des capacités* » (2000, p. 99). Cet échantillonnage consiste en la sélection d'items ou de situations constituant un échantillon suffisamment représentatif des contenus (connaissances, capacités ou compétences) ciblés. Outre la représentativité des éléments retenus dans les épreuves d'évaluation, De Landsheere insiste sur la nécessité d'intégrer les aptitudes intellectuelles supérieures pour assurer la validité de ces épreuves. Mais, sur quoi cette validité est-elle fondée ?

Nous savons depuis De Landsheere que la validité du contenu s'appuie sur des jugements. En effet, les juges - qui sont expérimentés dans le domaine cible – jugeront par concertation de la validité des questions proposées dans une épreuve ainsi que celle de la distribution des notes.

Toutes les définitions de ce type de validité sont unanimes sur la nécessité d'une « *congruence entre les items et les objectifs pédagogiques* » (Morissette & Laurencelle, 1993, p. 282), un élément que nous prenons en considération lors de la conception des situations que nous proposerons dans notre recherche. Cependant, nous retrouvons dans l'article de Fortin (1994), qui s'intitule « Propriétés méthodologiques des instruments de mesure (Fidélité – Validité », des informations sur ce type qui ne se réduisent pas à sa définition. En effet, l'auteur repère deux formes<sup>37</sup> de ce type de validité : la validité nominale et la validité par consensus. S'agissant de la validité nominale, elle consiste en un « *processus par lequel un expert juge de la validité d'un instrument de mesure en ce qu'il semble relié au concept d'intérêt* » (Fortin, 1994, p. 60). Quant à la validité par consensus, il s'agit du « *processus par lequel un panel d'experts juge de la validité d'un instrument de mesure* » (Fortin, 1994, p. 61). Ces deux formes, qui rejoignent les deux éléments abordés par Legendre (1985) dans sa définition de la validité du contenu : le lien des items avec le programme et leur représentativité, s'appuient sur le jugement des pairs (ou experts). En effet, la validité nominale telle que la définition de Fortin (1994) mobilise un expert tandis que la validité par consensus est fondée sur la concertation d'un groupe d'experts.

Nous tenons à préciser que ce type (la validité du contenu) est quasi-présent dans toutes les typologies proposées ou reprises dans différents ouvrages.

## **2.2. La validité prédictive**

Cette validité, dite également de critère, concomitante et critérielle, « *prédomine jusque dans les années 1950* » (André, Loye & Laurencelle, 2015, p. 132) au détriment de la validité manifeste. Néanmoins, Fortin (1994) ne rejoint pas cette équivalence entre les termes cités étant donné que, selon lui, la validité de critère correspond à un type qui englobe deux formes : la validité concomitante et la validité prédictive. Avant d'expliquer la classification établie par Fortin, nous définissons la validité prédictive en nous appuyant sur quelques études.

---

<sup>37</sup> Contrairement à certains auteurs qui emploient indifféremment les deux termes types et formes, une position que nous avons adoptée dans notre étude, Fortin (1994) distingue type et forme dans la mesure où le terme forme est réservé dans certaines parties de son article pour désigner, selon lui, les classes d'un même type.

Selon Demeuse et Henry, la validité prédictive repose sur la possibilité de « *prédire les résultats qu'obtiendront les sujets à d'autres tests ou à un instrument quelconque* » (2004, p. 163). Ainsi présenté, ce type assure la prédiction des résultats que les apprenants obtiendront dans de futurs tests en tenant compte de leurs performances manifestées dans le premier test. En effet, en évaluant la performance des apprenants dans une situation, nous repérons les critères maîtrisés, ce qui prédit les possibilités de leur réussite dans d'autres situations. Dans une conception qui rejoint celle de Demeuse et Henry, Fortin définit cette forme comme le « *degré de corrélation entre une mesure du concept et une mesure ultérieure du même concept* » (1994, p. 61), c'est-à-dire une comparaison des scores obtenus par les mêmes apprenants dans des tests évaluant les mêmes aptitudes à des moments différents. Contrairement à cette forme, la validité concomitante correspond à la corrélation entre les scores obtenus par les mêmes apprenants pour évaluer les mêmes aptitudes, au même moment, mais en recourant à des instruments variés, vu que la validité de critère est, selon Fortin (1994), relative à la corrélation entre des instruments pouvant mesurer le même phénomène.

Cette classification établie par Fortin (1994) ne s'écarte pas réellement de l'étude de Demeuse et Henry (2004) qui n'insistent pas sur les nuances entre types et formes. Il ne s'agit, pour nous, que d'une différence de points de vue, générée par la diversité des domaines étudiés<sup>38</sup>, sur ce type de validité : le point de vue de Demeuse et Henry (2004) insiste sur l'objectif alors que celui de Fortin (1994) explicite notamment les procédures. Etant donné que notre étude s'inscrit dans un contexte éducatif, nous rejoignons le point de vue de Demeuse et Henry qui met l'objectif de prédire les résultats dans d'autres instruments au premier plan.

De Landsheere (2000), qui rejoint Demeuse et Henry (2004), affirme la difficulté de prouver réellement la validité prédictive. En effet, la maîtrise des critères dans une situation donnée et à un moment donné n'en garantit pas la réussite dans d'autres situations. D'ailleurs, la performance manifestée par les apprenants dans un contexte scolaire n'est pas une preuve de compétence dans des situations quotidiennes et sociales.

### **2.3. La validité du construit**

Bien que ce type ne figure pas dans la distinction établie par De Landsheere (2000), la plus ancienne parmi celles dont nous disposons, Tardif (1993) le considère comme le type le

---

<sup>38</sup> Les recherches de Demeuse et Henry sont ancrées dans le cadre éducatif tandis que celles de Fortin (1994) portent sur le domaine médical, d'où la différence des points de vue.

plus important. Désignée par validité de construct ou conceptuelle, elle émerge vers les années 1950 (André, Loye & Laurencelle, 2015).

Appelé également validité conceptuelle, ce type consiste en « *la détermination des composantes de chaque compétence et, surtout, de leurs nécessaires interactions pour que la compétence puisse se développer* » (Tardif, 1993, p. 7). Cependant, la décomposition de la compétence en plusieurs composantes ne doit pas occulter leurs interactions qui sont d'une importance majeure. Dans cet ordre d'idées, Tardif (1993) évoque *l'invalidité du construit* si nous évaluons les composantes séparément dans la mesure où la maîtrise de ces composantes ne signifie pas théoriquement que la compétence est acquise. Cette conception doit être évoquée vu que les pratiques évaluatives portent notamment sur des situations complexes dans le cadre de l'APC, actuellement en vigueur par de nombreux systèmes éducatifs. Il importe que les situations proposées en évaluation apportent des « *informations sur les métamorphoses de chaque compétence de l'élève en rendant explicites ses forces et ses faiblesses cognitives ainsi que métacognitives* » (Tardif, 1993, p. 07).

Par opposition à la conception de Tardif de cette validité, celle de Demeuse et Henry semble restrictive. En effet, la validité du construit y semble exclusive à certains domaines où les objets évalués ne sont plus concrets. A titre d'exemple, les auteurs évoquent la mesure de l'intelligence et de la personnalité. Ils affirment que « *la validité de construct repose sur une définition de l'objet d'évaluation lorsque celui-ci n'est pas matérialisable en termes de contenus énonçables, comme dans le cas de la validité de contenu* » (Demeuse & Henry, 2004, p. 164). Le recours à ce type de validité est alors conditionné par la nature de l'objet d'évaluation. En effet, dans le cas des contenus concrets et observables, il est nécessaire de recourir à la validité du contenu. Dans le cas inverse, lorsque les objets à évaluer ne sont pas directement observables, le recours à la validité de construit est nécessaire. En nous appuyant sur ces mêmes propos, nous repérons également un lien établi entre la validité de construit et la validité du contenu dans la mesure où seul l'objet d'évaluation détermine le type de validité dont il est question.

Les trois formes de validité que nous venons de lister sont les plus connues. Selon André, Loye et Laurencelle (2015), elles sont ciblées en 1954 dans la première version des « *Standards for educational and psychological tests* ». En plus de ces types, les deux types suivants que nous expliciterons sont extraits de la typologie établie par Tardif dans son article. Il s'agit de la validité écologique et déontologique.

#### **2.4. La validité écologique**

Tardif fournit la définition suivante de ce type de validité : « *le degré de ressemblance entre les conditions dans lesquelles se trouvent les élèves dans une situation d'évaluation et celles qui existent à l'extérieur de la classe et de l'institution scolaire* » (Tardif, 1993, p. 07). Ce type semble extrêmement important dans la mesure où le premier objectif des apprentissages scolaires est leur investissement dans des situations sociales authentiques. En assurant ce type de validité dans les instruments évaluatifs, le jugement de compétence émis serait valide.

L'intégration de ce type de validité est perceptible dans les situations évaluatives où les enseignants tentent de reproduire certaines conditions naturelles pour évaluer la capacité de leurs apprenants à investir leurs acquis. En effet, dans une pédagogie d'intégration, c'est le transfert des acquis dans des situations authentiques qui est visé. Évaluer cette capacité de transfert requiert la reproduction de certaines conditions pour pouvoir prédire les chances de réussite dans la vie réelle. Cet objectif de garantir l'authenticité des conditions de l'évaluation scolaire pour exclure certains biais, tels que le décalage entre les conditions scolaires et les conditions sociales, nuisant à une mesure effective des compétences est difficile à atteindre pour certaines considérations. La première est que la reproduction des conditions authentiques par simulation ne les concrétise aucunement dans l'univers scolaire et ne prépare pas pleinement les apprenants à tous les imprévus de l'univers réel. La seconde est liée aux objectifs de l'école qui prépare progressivement les apprenants à la vie sociale : en voulant importer les conditions réelles en évaluation, nous heurtons cet objectif. Ces éléments expliqueraient l'absence de ce type dans les travaux de certains auteurs (André, Loye & Laurencelle, 2015 ; Demeuse & Henry, 2014).

En dépit de la richesse des informations fournies par Tardif sur ce type de validité, l'auteur ne fournit pas des explications satisfaisantes sur la manière de l'intégrer dans les pratiques évaluatives conçues et introduites dans les milieux scolaires. Il serait même important de préciser lesquelles parmi les conditions sociales il faudrait intégrer dans les situations évaluatives.

#### **2.5. La validité déontologique**

Tout comme le type précédent, ce type de validité que nous développerons est évoqué par Tardif. Selon cet auteur, il concerne « *le degré d'exactitude avec lequel un instrument ou une modalité d'évaluation détermine les compétences développées* » (Tardif, 1993, p. 08). En respectant ce type de validité, les instruments évaluatifs retenus doivent révéler avec exactitude

la compétence de l'apprenant. Cette validité n'est pas souvent évoquée vu que l'exactitude des instruments de mesure est un critère réclamé dans toute forme de validité.

En plus de ces deux types que nous avons développés en nous appuyant sur les travaux de Tardif, Demeuse et Henry proposent deux autres types que nous expliquerons succinctement dans ce qui suit.

### **2.6. La validité concourante**

Ce type consiste à « *valider un nouveau test à partir d'un autre, déjà utilisé et validé* » (Demeuse & Henry, 2004, p. 165). Il contribue à l'obtention de tests équivalents et identiques. Ce type semble adapté aux cas des tests contenant plusieurs items. En effet, les tests élaborés dans le cadre de l'APC contiennent des situations dont l'équivalence est assurée par le paramétrage.

### **2.7. La validité incrémentale**

Cette forme de validité - rarement évoquée selon Demeuse et Henry (2004) - est abordée par certains auteurs tels que Bernier et Pietrulewicz (1997). Dans cette validité qui recourt à une série de tests, un test est jugé valide lorsqu'il contribue à la valeur prédictive des tests.

### **2.8. Et la fiabilité ?**

Selon André, Loye et Laurencelle (2015), « *la fidélité est souvent considérée comme étant une forme de validité* » (p. 132). Cette notion qui se rapporte à la constance des résultats des tests (Buckingham, 1921) est cependant présentée comme un des critères professionnels de l'évaluation. Vu que la fiabilité est déjà traitée dans le premier chapitre de notre travail, nous ne la développons pas à ce niveau. Si nous l'évoquons à ce niveau, c'est pour montrer le lien étroit que ce critère entretient avec celui de validité et qui est à l'origine de la conception de la fidélité en tant que forme de validité.

Les avantages des formes de validité sont indéniables. En premier lieu, elles ont contribué significativement à mieux éclaircir la notion de validité étant donné qu'elles permettent de dégager les principaux indicateurs d'un examen valide. En second lieu, elles ont permis de concevoir des techniques et des stratégies de validation pour juger de la validité des tests administrés. En dépit de ces avantages, certaines de ces formes telles que la validité écologique et déontologique présentent des problèmes d'application. En effet, la première qui requiert la ressemblance des conditions scolaires du déroulement des tests aux conditions sociales semble exagérée alors que la seconde qui suppose exactitude de la mesure est une propriété intrinsèque à toutes les formes de validité. Eu égard aux contraintes d'application

posées par certaines formes de validité, serait-il légitime de considérer que certaines formes sont supérieures à d'autres ?

Comparer les formes de validité ne convient pas étant donné que chacune des formes évoquées se rapporte à une dimension particulière de la validité. Mais, de toutes les formes, la validité du contenu semble plus importante car elle est la plus liée à la validité. En effet, la définition de la validité formulée dans la conception classique (capacité des instruments de mesure à évaluer réellement ce qui est évalué) correspond particulièrement à celle de la validité du contenu. Rappelons que notre travail porte sur la validité des instruments de mesure et de jugements que le système éducatif met en œuvre en évaluation du FLE.

### 3. Des conceptions de validité aux procédures de validation

Dans une étude sur la validité de l'évaluation sommative, il est nécessaire d'aborder les procédures de validation dans la mesure où la démonstration de la validité d'une épreuve -ou toute évaluation au sens large- requiert le recours à des stratégies appropriées (Noël-Jothy, Sampsonis, 2006). Bien que les deux concepts (validité et validation) soient distincts, certains auteurs tels que Borsboom, Mellenbergh et Van Heerden (2004) mentionnent leur confusion manifeste dans quelques travaux sur la validité qui ont traité, à tort, la validité et la validation comme deux concepts interchangeables.

En effet, selon ces auteurs, « *la validité est une propriété, tandis que la validation est une activité* »<sup>39</sup>(Borsboom, Mellenbergh & Van Heerden, 2004, p. 1063). Par opposition à la validation qui comporte plusieurs procédures pour aboutir à un jugement fiable, la validité est une caractéristique affirmant que le test ainsi que ses résultats reflètent la ou les compétence(s) mesurées. Dans cet ordre d'idées, Noël-Jothy et Sampsonis proposent, dans leur glossaire, une définition de la notion de validation qui insiste sur son aspect actif. Ainsi, elle est perçue comme le « *processus qui consiste à prouver la validité d'un test, par lequel on rassemble des preuves pour étayer les conclusions données par les résultats des tests* » (2006, p. 116). Dans la même lignée, Borsboom, Mellenbergh et Van Heerden estiment que « *la validation est le genre d'activité que les chercheurs entreprennent pour savoir si un test a la propriété de validité* »<sup>40</sup> (2004, p. 1063).

---

<sup>39</sup>Traduction de « *validity is a property, whereas validation is an activity* » (Borsboom, Mellenbergh & Van Heerden, 2004, p. 1063).

<sup>40</sup>Traduction de « *validation is the kind of activity researchers undertake to find out whether a test has the property of validity* » (Borsboom, Mellenbergh & Van Heerden, 2004, p. 1063).

Nous constatons que dans la première définition de la validation, les auteurs associent ce concept à une argumentation visant à soutenir, par des épreuves appropriées, les résultats des tests ; alors que dans la seconde, le concept est plutôt lié aux activités soutenant la validité du test. Ceci indique une certaine divergence des perspectives au niveau de la notion de validité qui ne nuit aucunement à la distinction établie entre celle-ci et la validation.

Dans ce qui suit, nous aborderons les diverses procédures de validation. Nous commencerons par celles qui sont centrées sur les formes de validité que nous avons énumérées pour passer aux modèles, plus structurés et théorisés, qui constituent actuellement des références en évaluation.

### **3.1. Des formes de validité aux procédures de validation**

Les divers types de validité que nous venons d'expliquer ont donné lieu à plusieurs stratégies de validation des instruments de mesure. Ainsi, André, Loye et Laurencelle affirment que « *toutes les démarches apparentées aux formes de validité sont des procédures pour la validation* » (2015, p. 140). Dérivées de formes de validité, les stratégies de validation suivantes, reprises par Demeuse et Henry (2004), trouvent leurs origines dans les travaux de G. De Landsheere (1982).

#### **3.1.1. Pour la validité prédictive**

Purement empirique, cette validation -qui n'est pas facile à établir- est basée sur le contrôle des variables. Cependant, ceux-ci ne sont pas toujours des indices fiables pour la prédiction des divers phénomènes, d'où le recours aux critères dont le repérage n'est pas aisé en sciences humaines (Henry & Demeuse, 2004). S'agissant du repérage de ces critères, De Landsheere affirme que « *dans le monde physique, on dispose généralement de critères sans ambiguïté (...) En éducation, des critères aussi clairs sont rares, du moins si l'on se contente pas de signes finalement peu représentatifs de la réalité* » (2000, p. 201). La difficulté d'établir des critères clairs a conduit au recours aux prédicteurs qui consistent en des éléments ayant une influence démontrée sur le trait à prédire.

#### **3.1.2. Pour la validité du contenu**

Selon De Landsheere, « *la validité du contenu dépend de la qualité et de l'adéquation de l'échantillonnage des connaissances et des capacités* » (2000, p. 199). Pour que le test soit validé, il est nécessaire d'étudier la corrélation entre les items élaborés -qui constituent l'échantillon retenu- et le programme d'études (des connaissances, capacités, compétences). La qualité de cet échantillon dépend en grande partie du volume des contenus du programme sur



lesquels le test porte. Dans un test qui porte sur un contenu restreint, les questions posées touchent à tout ce qui est appris, d'où la représentativité de ces questions. Dans le cas du contenu volumineux, un choix judicieux s'impose pour la sélection des éléments à constituer les items du test. Ces éléments doivent révéler les compétences des évalués. En plus de leur représentativité, les items retenus doivent appartenir à divers niveaux taxonomiques (mémoire, compréhension, application, analyse, synthèse, évaluation). De Landsheere (2000) insiste sur ce second critère pour la validité du contenu.

Ainsi, la validation du contenu des examens s'opère par l'étude de la représentativité des questions et de leur appartenance à divers niveaux cognitifs.

### **3.1.3. Pour la validité du construit**

Pour étudier la validité du construit d'un test, nous mesurons à quel degré celui-ci recouvre les comportements qui lui sont attribués (Demeuse & Henry, 2004). Pour ce faire, *l'analyse factorielle*<sup>41</sup> est souvent présentée comme la stratégie la plus appropriée. Cependant, cette « démarche décrite ne donne évidemment pas de garanties absolues » (Demeuse & Henry, 2004, p. 168).

### **3.1.4. Pour la validité concourante**

Le dernier type de validité que les auteurs évoquent dans leurs procédures de validation est la validité concourante. Dans cette procédure, la validité du construit joue un rôle primordial : quand un test est validé, il devient une référence pour la validation de nouveaux tests. Parfois, il est même adapté de façon à répondre à de nouveaux besoins ou aux exigences de nouveaux contextes. En vue d'évaluer la validité concourante d'un test, une étude corrélative entre les résultats obtenus dans les deux tests par la même personne s'impose.

Ces démarches - traitées par Demeuse et Henry (2004) dans leur étude séparément des formes de validité - y sont étroitement liées : chaque forme est à l'origine d'une validation. En dépit de leurs avantages, elles ne sont pas suffisamment structurées pour devenir des modèles de large usage comme c'est le cas de la structuration argumentative de Kane (2006)... Ces démarches de validation inspirées des formes de validité suscitent un certain nombre de questionnements, ce qui nécessite des précisions sur la ou les démarches à retenir absolument : faut-il valider un test à partir d'une seule procédure ou devrait-on le faire en intégrant toutes les

---

<sup>41</sup> L'analyse factorielle consiste à étudier les facteurs liés à des données précises. Appliquée aux items d'un test, elle « permet de juger si le contenu du test est homogène, ou si les différents items mesurent des capacités largement distinctes » (Bernier, 1952, p. 649).

procédures ? Autrement dit, la validation d'un test pourrait-elle être déclarée à partir d'une seule procédure ?

### 3.2. Les modèles de validation

Nous avons cité ci-haut deux modèles incontournables de validation en évaluation scolaire auxquels les recherches ont abouti : la *structure argumentative* de Kane (2006) et the *Evidence-Centered Design* (2003), conçu par Mislevy et ses collaborateurs.

Ces modèles trouvent leurs origines dans les travaux de Messick (1989, 1994) où l'évaluation des connaissances était présentée comme « *un cas particulier de raisonnement de la preuve* »<sup>42</sup>(Mislevy, Almond & Lukas, 2003, p. 3), d'où la conception de modèles fondés sur des arguments et des preuves pour la validation des épreuves. En plus des travaux de Messick, ceux de Toulmin (1958), qui a développé une argumentation de développement de tâches basée sur des preuves, ont eu un apport incontestable dans la conception des deux modèles.

#### 3.2.1. Les travaux de Messick (1994)

Pour la conception d'une évaluation valide basée sur des preuves, Messick (1994) a explicité la démarche à suivre dans la citation suivante que nous empruntons à Mislevy, Almond et Lukas :

Une approche centrée sur la construction [la conception de l'évaluation] commencerait par demander quel ensemble de connaissances, compétences et autres attributs devraient être évalués (...). Ensuite, quels comportements ou performances devraient révéler ces contenus (connaissances et compétences) et quelles tâches ou situations devraient susciter ces comportements ? Ainsi, la nature des contenus guide la sélection ou construction de tâches pertinentes<sup>43</sup>(Messick, 1994, p. 14, cité dans Mislevy, Almond & Lukas, 2003, p. 3).

Dans cette citation, Messick établit plusieurs moments à respecter lors de la conception des évaluations. Le premier consiste à définir les connaissances et les compétences sur lesquelles doit porter l'évaluation, et ce conformément aux objectifs scolaires ciblés. En effet, la définition du trait qui constitue un échantillon du programme d'étude est importante dans la

---

<sup>42</sup> Traduction de « *An assessment is a special case of evidentiary reasoning* » (Mislevy, Almond & Lukas, 2003, p. 3).

<sup>43</sup> Traduction de « *A construct-centered approach [to assessment design] would begin by asking what complex of knowledge, skills, or other attribute should be assessed (...). Next, what behaviors or performances should reveal those constructs, and what tasks or situations should elicit those behaviors? Thus, the nature of the construct guides the selection or construction of relevant tasks* » (Messick, 1994, p. 14, cité dans Mislevy, Almond & Lukas, 2003, p. 3).

mesure où nous ne pouvons pas évaluer tous les contenus du programme d'études. Seul un échantillon représentatif est retenu en tant qu'objet d'évaluation... Une fois les objets à évaluer sont déterminés, il est nécessaire de préciser les manifestations observables concrètement pour certifier la maîtrise de ces contenus évalués.

Outre la sélection des contenus évalués, les contextes où ceux-ci sont intégrés doivent être bien étudiés pour élaborer par la suite des situations appropriées qui en permettent l'évaluation. Ainsi, des tâches ou des activités évaluatives permettant la manifestation du trait évalué sont sélectionnées. Mentionnons que les activités évaluatives sont déterminées par la nature des contenus : les connaissances sont évaluées par des items alors que les compétences sont mesurées à travers des situations complexes.

Suite à ce moment de conception, il est important d'établir l'argument ou les preuves de cette compétence évaluée. Cet argument se concrétise dans des éléments observables dans la production de l'apprenant, à partir desquels des critères sont établis.

La validation de Messick dont nous venons d'exposer les principales phases « *est un modèle souvent privilégié dans le cadre de la validation des questionnaires d'intérêts, impliquant une extension de la validité aux conditions d'utilisation de l'instrument et à ses conséquences* » (André, Loye & Laurencelle, 2015, p. 142). D'ailleurs, quelques auteurs l'ont investi dans leurs études en vue de vérifier la validité des épreuves étudiées. Parmi ces auteurs, André, Loye et Laurencelle citent Hébert (2013), Voss, Kunter et Baumert (2011).

### **3.2.2. Les travaux de Toulmin (1958)**

Toulmin a recouru aux arguments pour soutenir la validité des jugements et des affirmations. Bien que sa théorie ne soit pas exclusive à l'éducation, elle est appliquée en éducation pour tester la validité des épreuves.

Selon cet auteur, un argument solide doit être soutenu par des preuves quantitatives ou qualitatives nécessaires qui relèvent tout ce qui est perceptible et par-là même évaluable. A ce niveau de notre travail, nous devons mentionner que nous ne disposons pas de beaucoup d'informations à fournir sur la théorie de Toulmin et ses applications dans le domaine de l'évaluation scolaire.

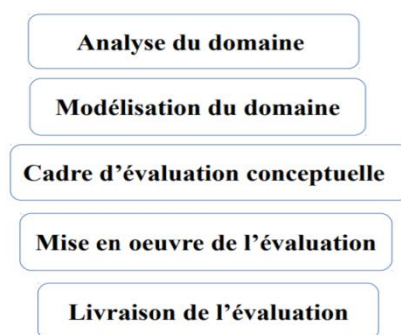
Les travaux de Messick et de Toulmin ont eu un apport incontestable sur les recherches en évaluation scolaire vu que plusieurs modèles centrés sur les arguments s'en sont inspirés. Dans les éléments qui suivent, nous aborderons les deux modèles que nous avons précédemment évoqués (la structure argumentative de Kane et Evidence-Centered Design de

Mislevy). De même, un troisième modèle de validation proposé De Ketele et Gérard, plus adapté aux nouvelles approches introduites telles que l'APC, sera traité.

### 3.2.3. The Evidence-Centered Design (2003)

Selon Mislevy et Riconscente, Evidence-Centered Design (conception d'évaluation basée sur les preuves), souvent abrégé en ECD, a été conçu pour « *explicitier, et pour fournir des outils pour, construire des arguments d'évaluation*<sup>44</sup> »<sup>45</sup> (2005, p. 1). Ces arguments contribuent d'une part à analyser les évaluations appliquées, et d'autre part à concevoir des évaluations valides.

Le modèle de Mislevy (ECD) s'organise en plusieurs couches. Nous les expliciterons en nous appuyant sur la représentation conçue par les auteurs précédemment cités.



**Figure 2 : Les niveaux du modèle ECD (Evidence-Centered Design)**

#### - Analyse du domaine

Cette première couche du modèle de Mislevy vise la collecte des informations nécessaires relatives au domaine cible. Parmi ces informations, Mislevy et Riconscente listent : le contenu (les connaissances), les concepts, la terminologie, les outils, les stratégies... Les informations recueillies englobent également les situations ou les contextes d'exploitation de ces connaissances ainsi que les stratégies susceptibles d'être mobilisées par les apprenants pour les résoudre.

La collecte de ces informations sert à la sélection des tâches appropriées. En effet, les tâches à proposer, qui portent sur des situations variées susceptibles d'être résolues de plusieurs manières, doivent révéler les ressources et les stratégies mobilisées par des apprenants. C'est à travers un bon investissement des savoirs, savoir-faire et stratégies dans les contextes adéquats

<sup>44</sup> Nous mentionnons que les mots (evaluation) et (assesment) qui renvoient au terme français (évaluation) ont des acceptions différentes.

<sup>45</sup> Traduction de « *explicit, and to provide tools for, building assessment arguments* » (Mislevy & Riconscente, 2005, p. 1).

que la compétence est prouvée, étant donné que celle-ci ne peut être mesurée qu'à travers des preuves repérées dans des productions concrètes.

Bien que ce niveau du modèle ECD ne se rapporte pas à la conception des outils et des activités évaluatifs, il jouit d'une importance incontestable pour la validité de l'évaluation. Une analyse judicieuse du contenu et du domaine cible contribue à orienter la conception vers des situations et des tâches appropriées. De même, elle assure la sélection d'indicateurs et de critères pour évaluer la compétence escomptée.

- Modélisation du domaine

A ce niveau de la conception d'évaluation basée sur les preuves, il s'agit de « *construire un schéma préliminaire de l'argument d'évaluation comme cadre général réutilisable pour une famille d'évaluations* »<sup>46</sup>(Almond, Mislevy, Steinberg, Yan & Williamson, 2015, p. 23). Nous évoquons un schéma préliminaire vu les modifications qui seront apportées dans les couches suivantes. En effet, ce schéma constitue une esquisse de l'argument d'évaluation qui sera régulée. Celui-ci consiste au lien établi entre les stratégies à mobiliser par les apprenants et les inférences existantes (les stratégies qui devraient être mobilisées).

Ainsi conçu, cet argument de compétence confronte les observations sur les actions des apprenants dans diverses situations aux actions qu'ils savent ou peuvent faire. Cet écart entre ce qui est manifesté par les apprenants et ce qui doit être fait peut être décelé à travers les situations-problèmes proposées.

Afin de collecter des preuves sur la compétence évaluée, il est important de concevoir des situations qui peuvent être résolus de diverses stratégies, ce qui permet de juger la capacité de sélectionner la méthode appropriée pour résoudre le problème posé.

Nous mentionnons que cette deuxième phase revêt un aspect descriptif : les traits généraux de l'évaluation sont abordés sans en évoquer les détails sur lesquels la couche suivante est axée.

- Cadre d'évaluation conceptuelle

Comme nous l'avons mentionné, cette couche permet de mieux traiter l'argument ébauché dans les deux premières couches. Ainsi, de la forme descriptive de la deuxième couche,

---

<sup>46</sup> Traduction de « *building a preliminary sketch of the assessment argument as a general, reusable framework for a family of possible assessments* » (Almond, Mislevy, Steinberg, Yan & Williamson, 2015, p. 23).

nous passons aux détails de la conception qui concernent de nombreux éléments tels que les procédures de la notation.

A ce niveau, nous rappelons que la compétence des apprenants ne peut être directement mesurée, ce qui conduit à une sélection soignée des tâches qui en permettent la manifestation. Ce n'est qu'à travers les productions des apprenants que les preuves de la compétence sont détectées.

Dans ce cadre de la conception d'évaluation centrée sur les preuves, Mislevy et Riconscente (2015) notent deux éléments importants qui doivent être pris en compte :

- Le composant d'évaluation

Selon Mislevy et Riconscente, le composant d'évaluation « *concerne les qualités des travaux produits par les apprenants : qualité, précision, élégance, stratégies utilisée* »<sup>47</sup> (Mislevy & Riconscente, 2006, p. 77). C'est à travers ces aspects saillants directement observables dans les productions réalisées et directement observables que les critères de notation sont déterminés.

La définition de ces critères est d'une importance majeure dans la mesure où elle contribue à cerner le travail réellement effectué par l'apprenant, ce qui rend aisé la procédure de mesure. Celle-ci sera complétée par un *mécanisme*, déterminant le degré de corrélation entre ce qui est réalisé (ou observé) et ce qui doit être réalisé, qui sera expliqué dans l'élément suivant :

- Le modèle de mesure

Ce mécanisme est un outil qui détermine le degré de corrélation entre ce qui est fait et ce qui doit être fait. De cette manière, il permet de mesurer l'écart entre la compétence réelle et la compétence manifestée.

S'agissant de la troisième couche de l'ECD, nous devons souligner que plus que le choix des tâches appropriées, il est important d'insister sur le rôle des conditions de mesure dans l'obtention de données réalistes sur la compétence des apprenants.

- La mise en œuvre de l'évaluation

---

<sup>47</sup>Traduction de « *Concerns the qualities of the Work Products students have produced—e.g., quality, accuracy, elegance, strategy used* » (Mislevy & Riconscente, 2006, p. 77).

A ce niveau, les éléments décrits dans les couches précédentes sont élaborés : la conception des tâches d'évaluation, l'ajustement des modalités de notation...

Le concepteur élabore ces tâches en s'appuyant sur un modèle. En effet, bien qu'elles varient dans certaines caractéristiques, elles ont en commun l'argument d'évaluation. Nous pouvons évoquer dans ce contexte *la famille de tâches*.

Il est important de mentionner que certaines décisions prises au cours de cette phase peuvent être ajoutées dans la couche qui la suit.

- Livraison de l'évaluation

A travers les couches suivantes, nous pouvons déceler toutes les informations sur les contenus relatifs au domaine cible, les connaissances et les compétences à évaluer, les procédures de cette évaluation grâce au repérage des aspects saillants déterminant des preuves de compétence... Ainsi, l'argument de preuve est défini.

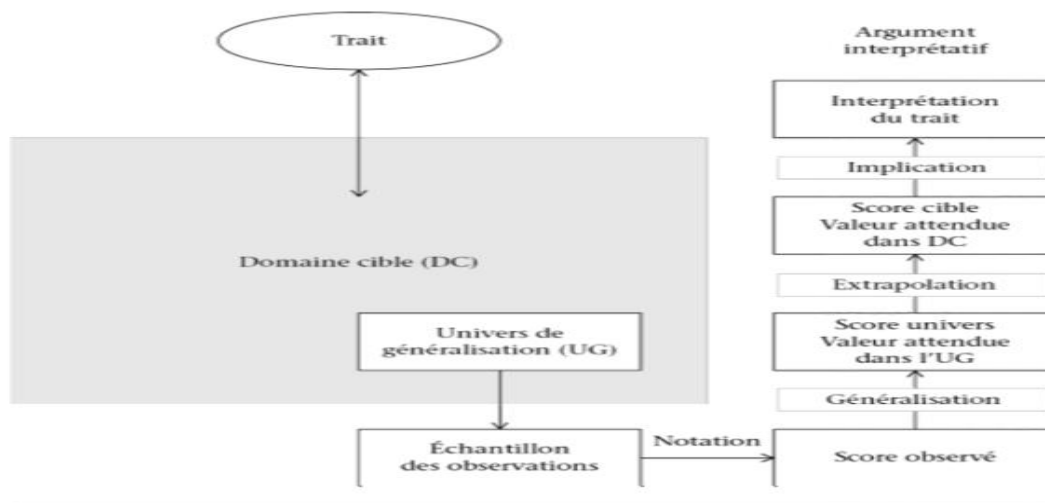
Cependant, toutes ces informations ne peuvent fournir des renseignements satisfaisants sur les stratégies auxquelles les apprenants recourent dans les divers contextes. Ce sont les interactions apprenant-tâche qui peuvent révéler des éléments nouveaux et utiles.

En somme, le modèle de Mislevy et ses collaborateurs a donné des renseignements précieux pour valider des évaluations déjà conçues ou en concevoir d'autres valides.

### **3.2.4. Structuration de l'argumentation interprétative (Kane, 2006)**

Élaboré en 2006, le modèle de validité argumentative de Kane constitue une référence dans l'analyse et l'élaboration des tests d'évaluation. Sa structuration est intégrée dans plusieurs ouvrages sur la validité de l'évaluation avec des adaptations pour répondre à des besoins particuliers : *Des mécanismes pour assurer la validité de l'interprétation de la mesure en éducation*, *Mesure et évaluation des compétences en éducation médicale*, *Manuel pour l'élaboration et la passation de tests et d'examens de langue...* Elle est adaptée pour être adoptée pour la validation des tests conçus ou la conception d'évaluations sommatives valides. Parmi ces ouvrages, le premier constitue à nos yeux une référence incontournable car il comporte des explications riches sur chacune des inférences de l'argument interprétatif.

En dépit de ces adaptations, quatre inférences de l'argumentation interprétative sont souvent retenues : la notation, la généralisation, l'extrapolation et l'implication.



**Figure 1: Structuration argumentative de Kane (Paquette-Côté, 2000, p. 13)**

Nous les expliciterons dans ce qui suit :

### 3.2.4.1. La notation

Cette inférence, qui correspond au premier niveau de l'argument interprétatif de Kane, consiste au passage de la performance observée au score. Elle comprend plusieurs éléments auxquels il faut tenir compte :

- Passation et notation appropriées

Le premier élément est la passation et la notation qui doivent être appropriées. Tel qu'il est employé par Karine Paquette-Côté et Gille Raïche (2013), le mot *approprié* semble important à préciser car il requiert une homogénéisation garante d'une certaine équité en évaluation. Cette condition exige une concertation au sein du personnel « *au sujet des objectifs, des contenus, des modes d'évaluation, des exigences et des seuils de réussite des cours et des programmes* » (Paquette-Côté & Raïche, 2013, p. 14).

Cette concertation des enseignants est d'autant plus importante car elle est requise dans toutes les phases sur lesquelles s'étendent la conception, la passation et la notation en évaluation sommative. En effet, elle débute par la sélection des objectifs d'enseignement qui doivent être unifiés, entraînant ainsi une homogénéisation des contenus proposés et nécessairement des programmes élaborés. Ces objectifs ne déterminent pas seulement les contenus enseignés mais ils sont décisifs dans le choix des items ou activités évaluatifs. Étant donné que toute évaluation porte sur ce qui a été enseigné, il devient aisé de prévoir et d'élaborer un sujet commun, à tous



les apprenants du même niveau, comprenant des items, un barème, des critères ou des grilles de correction identiques.

L'intégration des enseignants dès ce premier moment requiert une importance majeure dans la mesure où elle permet un échange d'expériences, ce qui aboutit à l'enrichissement des modalités et des outils évaluatifs intégrés. De même, elle permet certainement d'éviter un nombre important d'erreurs qui pourraient parvenir de l'exclusion de certains enseignants de cette opération. En effet, leur exclusion aboutit à une évaluation qui ne répond pas aux critères de validité et ne respecte pas l'équité et l'égalité entre les apprenants.

Ce premier élément correspond alors à la planification. Il consiste à établir les règles de passation et de notation qui prendraient en considération les opinions des différents acteurs.

- Application de ce qui est planifié

Dans la première inférence de la notation, le deuxième élément correspond à l'application ce qui a été planifié. A ce niveau, il convient de préciser qu'une application réussie est conditionnée par une planification judicieuse. En effet, la concertation des enseignants et leurs discussions lors de la planification et l'établissement des règles contribuent énormément à éluder les problèmes qui pourraient surgir au moment de l'application.

Dans la mesure où les règles sont établies en intégrant le corps-enseignant, leur application aurait plus de chances de réussir. Ce qui nous conduit à insister sur l'importance de la concertation de tous les acteurs et de la prise en compte de leurs avis et de leurs idées pour garantir une application réussie. Dans cet ordre d'idées, Karine Paquette-Côté et Gilles Raïche estiment que, les enseignants ayant participé au premier moment de cette inférence (la notation), « *on peut supposer qu'ils auront plus tendance à les appliquer telles que spécifiées* » (2013, p. 15).

Selon les mêmes auteurs, le degré d'application des règles de passation et de notation antérieurement établies par les enseignants serait précisé dans un rapport détaillé. En effet, cet élément, tout comme le précédent, est une condition nécessaire pour la validité de tout le processus.

- Exclusion de biais

Le troisième élément de cette inférence de la notation concerne l'exclusion de biais qui concernent tous les éléments qui peuvent affecter les conditions de passation et de notation contribuant ainsi à fausser les résultats des examens.

Selon K. Paquette-Côté et G. Raïche, ces biais peuvent être liés « *aux caractéristiques de l'évaluateur ou de l'étudiant ou à des caractéristiques environnementales* » (2013, p. 16). Afin de les écarter, plusieurs moyens doivent être intégrés assurant par-là la validité de l'inférence en question et des résultats obtenus.

Tous ces moyens en dépit de leur diversité se rapportent au respect des règles de notation et de passation déjà établies. Ainsi, la conception de tests doit se faire en respectant les objectifs, les critères, les seuils de réussite ainsi que les exigences définis. Rappelons que ces éléments sont explicités dès le premier élément de cette inférence pour réduire les risques de subjectivité qui peuvent biaiser la correction et la notation. Outre les éléments déjà cités pour quêter l'objectivité des enseignants, le respect des grilles et des critères préétablis est un gage de réussite assurant une évaluation équitable et objective.

Si le premier biais évoqué est relatif à la subjectivité des enseignants, le deuxième est relatif aux facteurs externes qui assurent la validité de l'évaluation. En premier lieu, les objectifs visés doivent être respectés et minutieusement sélectionnés dans la mesure où leur nombre important empêche leur évaluation intégrale, sauf si nous procédons à un test de connaissances comportant un grand nombre d'items. C'est la raison pour laquelle les enseignants sont contraints à sélectionner un échantillon représentatif à évaluer dans le cadre d'activités évaluatives appropriées. Outre les objectifs échantillonnés, des échelles de mesure et de notation doivent être clairement établies. Nous constatons que les biais évoqués dans cette première catégorie sont conditionnés par le respect des règles de notation et de passation antérieurement établies.

La deuxième catégorie de biais se rapporte aux conditions de passation qui peuvent nuire à la validité des tests. Celles-ci sont également étudiées et explicitées pour permettre aux enseignants de juger le rendement effectif de leurs apprenants et ce à travers le choix de la durée de l'évaluation, du nombre d'examens à passer par jour, du moment opportun de la journée en tenant compte des particularités du groupe enseigné... Ces éléments réunis sous la *standardisation de l'environnement du test* exigent le contrôle « *des facteurs tels que l'heure de l'examen, le temps alloué, les conditions d'assise, la température, la luminosité, la longueur*

*des questions, la taille de l'écran ou la qualité d'impression* » (Dionne & Raïche, 2017, s.p.<sup>48</sup>). Tout comme la subjectivité des enseignants, ces éléments cités peuvent fausser les résultats de l'évaluation, d'où la nécessité de les contrôler notamment dans un contexte de classe.

Enfin, la troisième catégorie de biais concerne les apprenants. Ceux-ci se rapportent à certains facteurs psychologiques (peur, stress...) qui altèrent les résultats obtenus. Pour les écarter et permettre aux apprenants de manifester leurs compétences effectives dans les tests qu'ils passent, il est important de les préparer psychologiquement à la passation et de leur fournir un exemple de test avant de le passer étant donné que la nouveauté des modalités d'évaluation peut biaiser l'évaluation. Un second biais se rapporte à la tricherie qui altère à son tour les résultats sur lesquels s'appuie le jugement interprétatif. Parmi les moyens permettant d'éviter ce biais, nous pouvons citer la sélection des instruments de mesure qui permettent de juger les compétences des apprenants, réduisant ainsi la tricherie qui nuit à la validité de l'évaluation. Plusieurs outils sont cités par Dionne et Raïche (2017) pour s'assurer de l'absence de tricherie. Parmi ces outils, il est important de « *ne pas réutiliser des items d'un test pour un autre test si le participant a accès à des ressources externes entre ses deux passations* » (Dionne & Raïche, 2017, s.p.<sup>49</sup>).

Alors, étant donné que l'évaluation n'est valide que si elle est exempte des trois catégories de biais citées ci-dessus. Ceux-ci peuvent être évités par la préparation psychologique et pédagogique des apprenants, l'instauration des conditions nécessaires et la formation des enseignants à toutes les pratiques évaluatives.

Le quatrième élément dans l'inférence de la notation se rapporte aux échelles de mesure qui doivent être clairement établies pour obtenir des scores valides.

Notre réflexion sur la première inférence de la structuration argumentative élaborée par Kane (2016) a incorporé quelques éléments de preuve qui en appuient la validité. Nous trouvons cependant ces éléments réunis chez Dionne et Raïche (2017). Nous les regroupons dans les points suivants :

---

<sup>48</sup> Ce livre est consulté sur le site Google Books. Les pages consultables ne sont pas numérotées.

<sup>49</sup> Sans page dans la source (Google Books).

- La précision et la représentativité des items sélectionnés : ce critère se rapporte à la capacité des items, qui constituent l'*univers de généralisation*, de représenter le *trait*, qui constitue une entité du *domaine cible*<sup>50</sup>.
- La crédibilité et la faisabilité du test : ces deux critères sont indispensables à la validité de l'évaluation. Les auteurs Dionne et Raïche insistent notamment sur celui de la faisabilité qui correspond à la « *facilité de réalisation à tous les niveaux. La clarté des instructions, la présence d'exemples, la durée impartie à la tâche par rapport à sa longueur ou sa difficulté font partie de ce critère* » (Noël-Jothy & Sampsonis, 2006, p. 42), ce qui revient à la nécessité de respecter les conditions assurant la validité antérieurement évoquée.
- La clarté des grilles de mesure élaborées : cet élément exige des concepteurs de l'évaluation l'élaboration de grilles aux critères bien précis reflétant les objectifs. De même, celles-ci devraient refléter les points forts et les points faibles chez les apprenants (ce qui est maîtrisé et ce qui ne l'est pas) dans la mesure où leur caractère informatif ne doit pas être négligé.
- La cohérence des modalités de notation : les échelles de notation et les barèmes sont élaborés en tenant compte des critères établis et de leur distribution est relative à l'importance de chaque item dans la manifestation du trait évaluée.
- La qualité des évaluateurs : cet élément conditionne la conception de tests précis et valides, l'obtention de résultats reflétant les compétences des apprenants, l'exclusion de biais qui altèrent les données obtenues...

Ces éléments sont récapitulés par Dionne et Raïche sous forme d'un tableau qui comprend plusieurs interrogations permettant aux enseignants de déterminer le degré de validité de la notation, première inférence du jugement argumentatif, dont l'objectif est de « *s'intéresser à la méthode de construction et de notation des tests, et de vérifier que les données issues du test ont été collectées de façon appropriée* » (Dionne & Raïche, 2017, s.p.<sup>51</sup>).

---

<sup>50</sup> Les concepts en italique dans le texte sont intégrés selon les significations qui leur sont attribuées par Dionne et Raïche dans leur ouvrage (Mesure et évaluation des compétences en éducation médicale- Regards actuels et prospectifs).

<sup>51</sup> Page non indiquée dans la source consultée (Google Books).

### 3.2.4.2. La généralisation

Selon Paquette-Côté et Raïche, c'est « *la constitution de l'échantillon des observations qui assure la validité de l'inférence de généralisation qui, à son tour, permet le passage de l'interprétation du score observé au score univers* » (2013, p. 18). Deux éléments de preuves valident ce niveau d'inférence : la représentativité de l'échantillon et sa grandeur.

La représentativité de l'échantillon renvoie au lien qu'entretiennent les items élaborés à l'univers de référence, autrement dit le programme ou l'ensemble des cours. Afin de l'assurer, l'enseignant élabore des activités évaluatives ou des items conformément aux objectifs visés dans les divers cours. En effet, la sélection de « *tâches/items pertinents (reflet des objectifs)* » (Noel-Jothy & Sampsonis, 2006, p. 47) est une des caractéristiques d'une évaluation valide.

En plus de sa représentativité, la grandeur de l'échantillon est importante pour éviter de nombreuses erreurs. En effet, le nombre des items retenus doit être suffisant pour refléter les objectifs institutionnels ainsi que les compétences des évalués. Ainsi, certaines conditions peuvent-elles biaiser les résultats obtenus, ce qui empêche une évaluation exacte dans la mesure où l'erreur commise par l'apprenant pourrait être due à un facteur externe. En multipliant les items, les erreurs récurrentes seraient indicatrices d'une maîtrise insuffisante du trait évalué.

En dépit de son importance, ce deuxième élément de l'inférence de généralisation peut susciter quelques réserves étant donné que le nombre des items dépend d'autres facteurs tels que le temps imparti au test. De plus, il semble incompatible avec l'évaluation dans le cadre de l'APC qui privilégie les situations complexes comme activités évaluatives car elles fournissent des données utiles sur les stratégies des apprenants.

Selon Dionne et Raïche, les « *éléments de preuve soutenant l'inférence de généralisation concernent la représentativité des observations et la reproductibilité du test* » (2017, s.p.). Pour le premier élément, la représentativité des items par rapport à l'univers de généralisation est assurée par le respect des objectifs et la multiplicité des activités pour éviter les biais. Outre ces deux critères, la représentativité de l'échantillon est déterminée par la nature de l'univers. En effet, une distinction est établie entre l'*univers homogène* dans lequel le nombre des items requis peut être limité et l'*univers hétérogène* dont l'échantillon doit être large et varié pour garantir sa représentativité. Ainsi, la nature de l'univers de référence (homogène ou hétérogène) contribue à établir la nature des instruments de mesure que comprend le test ainsi que le nombre (restreint ou large) d'items ou situations retenus.

Le second élément évoqué par Dionne et Raïche dans leur ouvrage est la *reproductibilité* de l'échantillon. Ce concept désigne « *la probabilité d'obtenir des scores similaires si l'on utilise un nouvel échantillon de situations* » (Dionne & Raïche, 2017, s.p.). A ce niveau, les résultats obtenus doivent être constants quel que soit l'échantillon de situations que comprend le test. Cette constance qui est révélatrice de la représentativité des échantillons choisis aboutit à des résultats fiables dont l'interprétation serait valide.

Selon les mêmes auteurs, « *l'objectif de cette deuxième inférence est d'étudier dans quelle mesure il y a corrélation entre le score « observé » et le score « univers »<sup>52</sup> inconnu, mais correspondant théoriquement à l'ensemble de tous les items représentant le domaine d'intérêt* » (Dionne & Raïche, 2017, s.p.). Cette corrélation entre les deux scores se concrétise dans la représentativité de l'échantillon et sa reproductibilité, déjà expliqués, qui constituent les éléments de preuve de l'inférence de généralisation. Soulignons que la première preuve de cette inférence de généralisation, mentionnée par Paquette-Côté et Raïche (2013), ou représentativité, est identique à celle citée par Dionne et Raïche (2017). Cependant, la deuxième preuve ne constitue pas un objet de consensus entre les auteurs cités.

Tout comme la première inférence, la généralisation incorpore des éléments de preuve qui soutiennent la validité des tests et contribuent à leur conception. Mais, par opposition à la première inférence dont certains traits sont relatifs aux conditions psychométriques, celle-ci est particulièrement liée à la matière scientifique enseignée.

### **3.2.4.3. L'extrapolation**

L'objectif de ce troisième moment « *est d'étudier la corrélation entre le score univers et le score cible dans le domaine d'intérêt ; en d'autres termes, l'extrapolation nous emmène de l'univers de test vers le monde réel* » (Dionne & Raïche, 2017, s.p.). En effet, l'objectif de l'évaluation -qu'elle porte sur des connaissances ou sur des compétences- ne se limite pas à déterminer ses savoirs et savoir-faire, mais d'avoir des données sur la capacité de transférer et d'investir ce qui est appris dans les situations appropriées. C'est la valeur prédictive qui prime : à travers les items et les activités retenus, nous pouvons prédire la compétence de l'apprenant à agir dans le monde réel. Si nous nous mettons dans une optique d'évaluation de compétences, la compétence évaluée devient ainsi la prédiction de la performance à manifester.

---

<sup>52</sup>Entre guillemets dans le texte original.

A ce niveau de l'argumentation interprétative, « *le score est alors interprété en tenant compte du programme d'études et du domaine cible* » (Paquette-Côté & Raïche, 2013, p. 21). Ce qui permet d'émettre des jugements sur les compétences de l'évalué dans le domaine cible, plus développé que l'univers de généralisation.

Tout comme les deux inférences précédentes (la notation, la généralisation), celle de l'extrapolation repose sur deux éléments :

- Le lien entre le score univers et le score cible

Le programme d'études n'est qu'un élément qui est en lien avec un domaine plus large. Etroitement lié au programme d'études, le *score univers* est alors interprété de façon à refléter un aperçu sur les compétences des apprenants dans le domaine cible.

Il est important de mentionner que les cours dispensés par l'enseignant sont renforcés par des lectures supplémentaires permettant aux apprenants d'approfondir leurs connaissances dans le domaine cible. Ainsi, le contenu sélectionné qui est en lien avec ce domaine constitue un échantillon représentatif de celui-ci. Il en est de même pour les items retenus en évaluation, dont les résultats obtenus reflètent les résultats que les apprenants pourraient obtenir dans le domaine cible.

Rappelons que la représentativité des résultats est notamment conditionnée par le recours à des instruments de mesure appropriés.

- Contrôle des biais

Le deuxième facteur cité par Raïche, Ndinga et Meunier pour assurer la validité de l'extrapolation est le contrôle des erreurs susceptibles de la biaiser. Selon ces auteurs, « *la source d'erreur la plus souvent mentionnée est le niveau de langue parlée et écrite. En effet, l'évaluation des apprentissages d'un étudiant peut être biaisée s'il ne possède pas les compétences linguistiques requises par la méthode d'évaluation* » (Paquette-Côté & Raïche, 2013, p. 23). A ce niveau de l'argumentation interprétative, il est nécessaire de ne pas nuire à la validité de la mesure par des compétences auxiliaires. Ce cas-ci, évoqué par les chercheurs, est celui des disciplines autres que les langues. Il devient intéressant de mener une étude rigoureuse pour déterminer les biais qui conduisent à des erreurs altérant la validité de cet argument en évaluation des langues telles que les opinions manifestées des apprenants, le manque d'informations sur un sujet donné... Alors, les compétences auxiliaires et les connaissances des apprenants sur lesquelles l'évaluation d'une langue ne doit pas porter constitueraient des éléments qui biaisent la validité de ce niveau d'inférence.

Afin de soutenir cette inférence, Dionne et Raïche dans un tableau récapitulatif proposent des éléments qui orientent les questionnements des évaluateurs. S'agissant du premier élément du tableau, il renvoie à la réduction des biais qui altèrent les résultats des épreuves ainsi que leurs interprétations. Pour ce qui est du second, il se rapporte à la corrélation entre le test et les méthodes d'évaluation du même trait. Cette corrélation devient d'une utilité incontestable lorsque le test ne permet pas de recueillir des preuves suffisantes sur la compétence de l'apprenant.

Ayant abordé le concept « trait » à ce niveau, il nous semble nécessaire d'en préciser le sens selon la conception argumentative de Kane. En effet, celui-ci ne renvoie pas à un item particulier, mais il correspond à « *une disposition à agir d'une certaine façon en réponse à certains types de stimuli ou de tâches, dans certaines circonstances* » (Dionne & Raïche, 2017, s.p.). Inspirée des travaux de Kane, cette définition avancée par Dionne et Raïche présente le trait comme la compétence évaluée qui permet à l'apprenant de résoudre les diverses tâches ou activités que comporte le test.

Enfin, il nous semble important de déterminer que le score observé pour l'inférence de généralisation est le même que celui de l'inférence d'extrapolation. Le passage à l'extrapolation consiste à étendre l'interprétation à un domaine plus large que l'univers de généralisation, qui renvoie au programme ou aux cours du domaine cible englobant une multitude de contextes où le trait serait investi.

#### **3.2.4.4. L'implication**

La dernière inférence de la structure argumentative de validation de Kane renvoie à l'implication. A ce niveau de l'argumentation, « *il s'agit de considérer dans l'interprétation du score les contextes dans lesquels la ou les compétences évaluées sont impliquées, les contextes scolaires et professionnels, les contextes de mesure et du quotidien* » (Paquette-Côté & Raïche, 2013, p. 24). Cette prise en considération de la diversité des contextes aboutirait à une multitude de suggestions, d'affirmations et de propositions qui constituent les *implications* du trait. Ces implications concernent également les conditions que doit remplir tout apprenant pour être reconnu compétent dans un domaine particulier, les exceptions ou les conditions qui peuvent altérer le résultat (les biais tels que les compétences auxiliaires)...

Deux éléments entrent en jeu pour soutenir cette dernière inférence. En premier lieu, il est nécessaire que toutes les interprétations associées au trait ou à la compétence soient incluses. En effet, la compétence de l'apprenant est déterminée en intégrant toutes les interprétations



antérieurement formulées et en prenant en considération les contextes variés (scolaire, social...) où la compétence serait transposée.

En second lieu, Raïche, Ndinga et Meunier évoquent l'adaptation du score aux implications associées à la compétence où deux positions se présentent. Dans la première, chaque cours est clos par une évaluation, mais c'est l'évaluation finale qui porte sur un ensemble de cours qui permet de définir la compétence de l'apprenant. La deuxième position requiert la variation des activités aux contextes variés pour mesurer avec exactitude le trait. En effet, la compétence de l'apprenant ne peut être validée en s'appuyant sur une seule situation, d'où certaines règles appliquées en évaluation scolaires (telle la règle 2/3 selon laquelle la compétence n'est maîtrisée que si elle est réussie dans deux situations sur trois).

Les arguments avancés par Dionne et Raïche pour soutenir cette inférence « *concernent la crédibilité de la prise de décision et les conséquences de l'évaluation* » (2017, s.p.).

- La crédibilité de la prise de décision : cet élément est assuré par l'intégration de toutes les implications issues des inférences précédentes pour aboutir à un jugement valide (argumenté) et à une décision fiable (réussite ou échec).
- Les conséquences de l'évaluation : ce second élément de preuve est relatif à l'aspect utilitaire de l'évaluation qui doit aboutir à des données répondant aux exigences institutionnelles et sociales.

Inspirée des travaux de Messick (1989), le modèle de Kane (2006) a été reproduit et adapté par plusieurs auteurs. Il constitue une référence incontournable pour argumenter la validité de tests à travers plusieurs phases : les inférences. La validation de ces inférences constituant le modèle de Kane assure la validation du test et de ses résultats. Ainsi, toute faille au niveau de l'une des inférences explicitées nuit à la validité du jugement interprétatif formulé. Selon le *Manuel pour l'élaboration et la passation de tests et d'examens de langue*, « *si l'objectif final est que le test soit valide, chaque étape doit être finalisée de façon satisfaisante* » (Conseil de l'Europe, 2001, p. 16). Alors, il est nécessaire de valider toutes les inférences de la structure argumentative pour qu'une déclaration de validité soit formulée.

En dépit de des diverses adaptations et applications de ce modèle, il n'a pas échappé à certaines critiques que des auteurs ont soulevées. Nous les développerons dans ce qui suit.

### 3.2.4.5. Limites du modèle de Kane (2006)

Théoriquement, le modèle de Kane tient compte de toutes les phases de la conception et de l'interprétation des examens. Mais, plusieurs auteurs tels que De Ketele et Gérard (2005), Howe et Ménard (1993) mentionnent les limites de son application dans certains contextes.

Dans son travail sur l'argumentation de la validité des inférences en évaluation, Paquette-Côté reprend Howe et Ménard (1993) qui affirment que « *certaines pratiques d'évaluation des apprentissages dans le réseau collégial ne permettent pas la validation de l'argument interprétatif pour l'interprétation d'un trait parce qu'elles présentent des failles d'inférence* » (Paquette-Côté, 2000, p. 27). Ce résultat, auquel a abouti l'étude menée par les deux auteurs cités, a conduit Paquette-Côté à s'interroger sur la compatibilité de ce modèle avec la nouvelle approche actuellement appliquée : l'Approche Par les Compétences.

En effet, le trait est évalué à travers plusieurs items représentatifs de l'univers de généralisation que comprend le test. Cette évaluation par items serait plus compatible avec l'Approche Par les Objectifs (APO) où les connaissances constituent les objets évalués. Pour l'évaluation dans le cadre de l'Approche Par les Compétences (APC), les activités évaluatives préconisées consistent en des situations complexes qui permettent l'investissement des ressources.

Cette limite de la structure argumentative de Kane est explicitée par De Ketele et Dufays (2003) qui ont établi trois pratiques évaluatives en corrélation avec les approches appliquées à l'enseignement/apprentissage des langues :

- Dans l'optique des contenus, évaluer consiste à prélever un échantillon de contenus représentatif de l'univers de référence des contenus enseignés ;
- Dans l'optique des objectifs, évaluer consiste à prendre un échantillon représentatif d'objectifs spécifiques et opérationnels et à générer un échantillon de questions qui traduisent au mieux cet échantillon d'objectifs ;
- Dans l'optique des compétences, évaluer consiste à proposer une ou des situations complexes, appartenant à la famille de situations définie par la compétence, qui nécessitera (ont) de la part de l'élève une production elle-même complexe pour résoudre la situation. (De Ketele & Dufays, 2003, Cités dans De Ketele & Gérard, 2005, p. 2).

De cette citation, il est aisé d'extraire l'activité évaluative qui correspond à l'approche pédagogique appliquée. Ces diverses activités évaluatives sont antérieurement évoquées par De Ketele (1988) qui a proposé quatre niveaux d'items : les items de restitution, les items d'application, les items de résolution de problèmes et les items de réalisation.

En effet, évaluer dans le cadre de l'APC requiert le recours aux situations (items de résolution de problèmes) où les apprenants transfèrent ce qui est appris (les ressources). Cependant, cette évaluation n'est pas aussi simple à valider. Ainsi, parmi les principaux critères qui fondent la validité d'une épreuve, De Ketele et Dufays (2003) citent le caractère représentatif de l'échantillon.

Si cette représentativité des items assure la validité des épreuves conçues dans le cadre de l'Approche Par les Objectifs, il n'en est pas de même de l'APC dont les épreuves n'obéissent pas à la même logique. Dans ce contexte, De Ketele et Gérard affirment que « *les épreuves élaborées selon l'approche par les compétences risquent (...) de présenter des difficultés de validité et de fiabilité étant donné la taille de l'échantillon issu de l'univers de référence* » (2005, p. 6). En effet, une situation qui mobilise plusieurs ressources ne pourrait jamais incorporer tous les éléments de l'univers de référence. Les items alors peuvent exister dans le cadre de cette approche car les situations complexes proposées exigent l'investissement de plusieurs savoirs, savoir-être et savoir-être.

C'est la difficulté de valider les épreuves élaborées selon l'APC qui a conduit les De Ketele et Gérard à concevoir des outils de validation plus adaptés aux épreuves évaluatives conçues dans le cadre de cette nouvelle approche.

### **3.2.5. Validation des épreuves selon De Ketele et Gérard**

Les limites du modèle de Kane (2006), soulevées par De Ketele et Gérard, les ont amenés à proposer des validations plus appropriées à l'approche par les compétences, une approche actuellement appliquée dans plusieurs systèmes éducatifs. Ainsi, les auteurs envisagent les validations suivantes :

- Une validation a priori par recours aux juges,
- Une validation empirique externe,
- Une validation empirique interne.

Les deux derniers types de validation énumérés sont regroupés sous l'appellation validation *empirique a posteriori*. De cette manière, nous aboutissons à deux validations que nous développerons dans ce qui suit.

#### **3.2.5.1. La validation a priori par recours à des juges**

Il s'agit d'une technique classique de validation qui consiste en la vérification par des juges (experts ou enseignants ayant les connaissances requises sur le programme scolaire) de la corrélation entre les items retenus dans une épreuve et l'univers de référence (le programme

d'études). Dans le cadre de l'APC, les « *items sont les situations-problèmes concrètes qui sont proposées aux élèves et l'univers de référence est la famille de situations qui correspond à la compétence* » (De Ketele & Gérard, 2005, p. 7) évaluée. Il est utile de préciser en quoi consiste le concept famille de situations qui constitue l'univers de référence pour préciser en quoi consiste cette validation. Selon Miled Mohamed, il s'agit d'« *un ensemble de situations ayant un niveau de difficulté équivalent mais qui se rattachent à une compétence ou à un même OTI* », c'est-à-dire un Objectif Terminal d'Intégration (2005, p. 131), celui-ci étant relatif au profil de sortie d'un cycle.

Le concept de famille de situations étant précisé, il devient nécessaire de répondre à la question cruciale qui se pose à ce niveau : comment prouver la validité d'une situation-problème ?

Une situation d'intégration sélectionnée est déclarée valide lorsque les paramètres, précis et clairs, de la famille de situations y sont présents. De ce fait, il est nécessaire de « *disposer d'une description suffisamment précise de la famille de situations, c'est-à-dire d'une définition des paramètres qui la caractérisent* » (De Ketele & Gérard, 2005, p. 7). Etant donné que le paramétrage d'une situation est utile au processus de validation, les auteurs ont établi les divers paramètres (type de la situation, tâche attendue, critères de correction...) qui définissent les situations appartenant à une même famille. Nous ne nous étalons pas sur ces paramètres qui seront abordés dans le chapitre suivant.

Il est nécessaire que ces paramètres non-exhaustifs soient investis pour la validation des situations d'évaluation et de son appartenance à une famille de situations. Dans le cas où les paramètres des situations proposées sont déjà fournis, ils sont appliqués. Dans le cas échéant, les juges sont contraints à les concevoir.

Bien qu'elle soit conçue en réaction aux limites du modèle de Kane, cette validation de De Ketele et Gérard (2005) rejoint le modèle de Kane (2006) dans quelques éléments. Dans ce contexte, Paquette-Côté (2010) insiste sur la compatibilité de cette technique avec les inférences de la structure argumentative. En effet, elle affirme que cette technique est

Tout à fait compatible avec le modèle de validation des arguments interprétatifs proposés par Kane (2006) puisque des juges, experts du domaine cible, possèdent effectivement l'expérience nécessaire pour juger des items ou des situations-problèmes qui font effectivement partie du domaine cible (Paquette-côté, 2010, p. 30).

Cependant, le modèle de Kane est enrichi par l'intégration des principales notions (situation-problème, famille de situation...) liées à la nouvelle approche. De cette façon, la validation de De Ketele et Gérard (2005) constitue une actualisation du modèle de Kane pour une meilleure adaptation aux principes de l'APC.

### 3.2.5.2. Les validations a posteriori

Ces validations empiriques englobent deux volets : une validation interne qui vise « à vérifier l'unité conceptuelle de l'épreuve en elle-même » (De Ketele & Gérard, 2005, p. 09), et une validation externe qui « tend à montrer la convergence de l'épreuve avec d'autres épreuves similaires elles-mêmes validées ou à confronter des résultats avec d'autres épreuves » (De Ketele & Gérard, 2005, p. 09). Nous les expliquerons dans les éléments suivants.

#### - La validation empirique interne

Elle s'opère par deux techniques : l'analyse factorielle et le calcul des coefficients d'homogénéité.

Selon Cibois (2006), l'analyse factorielle est créée par Spearman<sup>53</sup> en 1904, pour mesurer l'intelligence. Elle a été généralisée, par la suite, à divers domaines en sciences. Elle « est régulièrement utilisée en sciences sociales et sert généralement à l'évaluation de la validité de construit d'instruments psychométriques » (Bourque, Poulin & Cleaver, 2006, p. 325). Intégrée dans le domaine de l'éducation, elle vise à valider les épreuves en fonction de leurs natures. En effet, dans le cas des épreuves qui évaluent une compétence particulière, nous étudions les corrélations des différents items avec le facteur principal extrait, ce qui contribue à une variance expliquée importante. Quant aux épreuves composites, il est nécessaire de vérifier si, outre le facteur principal extrait, nous trouvons des facteurs représentant les éléments variés constituant l'épreuve.

La deuxième technique, le calcul des coefficients, qui vise à valider la cohérence interne des instruments de mesure est régi par le même principe (lorsque l'épreuve évalue une seule dimension, nous vérifions si le coefficient obtenu est suffisamment élevé) ou à déterminer les items qui n'évaluent pas ladite dimension. Dans le cas des épreuves aux variables multiples, le nombre des coefficients calculés dépend des dimensions incluses.

---

<sup>53</sup> Charles Edward Spearman (1863-1945) est un psychologue anglais. Il est connu pour ses recherches sur l'intelligence et sur l'analyse factorielle.

Cependant, De Ketele et Gérard (2005) considèrent que les épreuves d'évaluation conçues dans le cadre de l'APC ne s'approprient pas à ces analyses corrélationnelles de l'analyse factorielle en raison de la complexité des situations qui intègrent plusieurs dimensions au caractère multidimensionnel. C'est la raison pour laquelle la validation des épreuves devrait s'appuyer sur des critères d'évaluation qui sont constants pour les situations appartenant à une même famille. Soulignons à ce niveau qu'une distinction s'établit entre *critère* et *indicateur*. En effet, le premier est commun à toutes les situations d'une même famille alors que le second est spécifique à une situation particulière.

La notion de critère, un concept-clé dans l'évaluation des compétences, est définie dans plusieurs ouvrages. Selon Roegiers, le critère de correction consiste en « *une qualité que doit respecter la production d'un élève* » (2005, p. 111). Ainsi, les critères deviennent les qualités de la production que les enseignants attendent de la leurs apprenants suite à la résolution d'une situation-problème. Dans le même ouvrage, l'auteur ajoute que le critère consiste en un point de vue adopté pour juger une performance évaluée. Cette deuxième définition qui rejoint la première suggère une certaine subjectivité des critères qui sont définis par le correcteur. Dans la mesure où une situation-problème comprend plusieurs dimensions évaluables, il est nécessaire de définir plusieurs critères. Leur nombre doit cependant être restreint car chaque critère se multiplie en indicateurs observables et spécifiques à la situation.

Selon De Ketele et Gérard, cette validation interne, basée sur les critères internes des situations, « *consistera à comparer diverses situations relatives à la même compétence : les élèves qui ont des scores élevés sur un des critères dans une situation devraient avoir des scores élevés sur le même critère dans une autre situation appartenant à la même famille, et inversement* » (2005, p. 10). Ainsi conçu, ce processus de validation requiert en premier lieu le paramétrage des situations et la confrontation de celles qui évaluent la même compétence, ce qui contribue à déterminer une famille de situations. Étant donné que les critères sont communs aux situations de la même famille, la réussite d'un critère dans une situation proposée en prédit la réussite dans les autres. En soumettant les élèves à des situations identiquement paramétrées, les critères maîtrisés par les apprenants sont réussis dans plusieurs occasions.

Il est utile de préciser qu'il est nécessaire d'identifier deux types de critères évoqués par De Ketele et Gérard (2005) en évaluation d'une compétence : le critère minimal et le critère de perfectionnement. Cette distinction est utile car elle conditionne la répartition des points sur l'échelle des notes. Il importe de préciser que le taux réservé aux critères minimaux est manifestement plus important que celui des critères de perfectionnement.

Tout comme la validation a priori par recours aux juges, la validation empirique interne est associée par Paquette-Côté, dans son mémoire, à la validité de l'inférence de la notation dans la structure argumentative de Kane (2006). Pour elle, « *à cette validation empirique interne peuvent être associés tous les arguments de validité de l'inférence de notation* » (Paquette-côté, 2010, p. 31). En effet, De Ketele et Gérard (2005) en citent plusieurs arguments tels que la comparaison des résultats, la distribution sur les situations, les corrélations entre les notes attribuées à un même critère...

La correspondance entre cette validation interne et le modèle de Kane a conduit Paquette-Côté (2010) à infirmer toute incompatibilité du modèle de Kane avec la nouvelle approche adoptée par de nombreux systèmes éducatifs (APC).

- La validation empirique externe

En plus des stratégies de validation interne explicitées ci-dessus, d'autres relevant de la validation externe peuvent être investies. Pour cette deuxième catégorie de stratégies, il est nécessaire de recourir à des critères externes impliqués dans le cadre d'études comparatives ou corrélationnelles.

Les études comparatives sont fondées sur une comparaison d'un groupe expérimental à un groupe contrôle. Il importe de préciser que la constitution du groupe contrôle doit être judicieusement étudiée. En effet, De Ketele et Gérard (2005) évoquent deux cas de figures :

- L'Approche Par les Compétences n'est pas généralisée : dans ce cas-ci, le groupe contrôle est extrait des établissements où l'approche n'est pas encore appliquée et les enseignants ne sont pas encore formés aux nouvelles pratiques correspondantes...
- L'Approche Par les Compétences est généralisée : dans ce second cas, le groupe contrôle est constitué d'apprenants qui n'ont pas appris à investir les ressources acquises dans la résolution de situations-problèmes.

Dans les deux cas, en proposant une situation identique aux deux groupes, « *la maîtrise de chacun des critères évaluant la compétence concernée devrait être supérieure dans le groupe expérimental* » (De Ketele & Gérard, 2005, p. 13).

Pour les études corrélatives, De Ketele et Gérard (2005) les fondent sur un seul groupe, contrairement aux études comparatives. Plusieurs situations sont proposées à ce groupe pour vérifier la maîtrise de la compétence et des critères qui lui sont associés. Les performances

manifestées dans les diverses situations permettent de valider les jugements émis étant donné la réussite des mêmes critères est un indicateur de la compétence du groupe évalué.

Sur ces deux techniques de validation externe, plusieurs questions se posent. En effet, les auteurs n'expliquent pas si ces techniques doivent être toutes investies pour la validation des épreuves et de leurs résultats.

En dépit de leur utilité, les diverses techniques de validation sont loin d'être exemptes de défaillances. En effet, un jugement de compétence soutenu par la réussite des critères semble infondé : un critère réussi dans une ou deux situations pourrait être raté dans d'autres situations. La vérification de la maîtrise du critère et sa généralisation à celle de la compétence sont biaisées en raison de la complexité des situations et leur multi-dimensionnalité, deux facteurs qui compliquent la prédiction de la réussite des critères. Dans ce contexte, De Ketele et Gérard affirment qu'« *il n'est pas sûr que la maîtrise de chacun de ces critères soit la preuve de la maîtrise de la compétence visée (...) parce que son évaluation est relative aux critères utilisés (leur nombre, leur opérationnalisation, leur statut)* » (2005, p. 14). Cette limite de la valorisation d'une compétence à travers la réussite répétitive des critères a conduit Paquette-Côté à estimer plus fonctionnel le modèle de Kane qui propose une démarche de validation complète englobant les diverses inférences dont la validité détermine celle de l'argument interprétatif en entier (Paquette-Côté, 2010).

Dans notre étude pratique, nous n'appliquons pas toutes les procédures de validation exposées tout au long de ce chapitre. Nous investissons celles qui permettent la validation des situations évaluatives conçues dans le cadre de l'APC, étant donné que toute proposition de dispositifs évaluatifs dans la présente étude doit être conforme aux exigences et aux principes de cette approche. Nous ne pouvons néanmoins nier l'apport considérable des autres modèles de validation qui fournissent des précisions précieuses sur les étapes à suivre au moment de l'élaboration des épreuves évaluatives. C'est le cas du modèle de Messick (1994), qui en dépit de son ancienneté, énumère les mêmes phases actuellement intégrées en élaboration des situations d'évaluation. Quant au modèle Evidence-Centered-Design (2003), il est aussi utile que le premier car nous y retrouvons des informations similaires. Pour la validation de Kane (2006), elle fournit aux enseignants-évaluateurs des précisions garantes de jugements valides fondés sur des inférences dont la validité conditionne la validité de tout le processus. Cependant, cette validation dont les inférences valident les résultats obtenus suite à la passation néglige la validité d'un élément crucial dans toute évaluation scolaire : les outils évaluatifs. En effet, en dépit de quelques remarques relatives à la représentativité des items ou situations du programme



scolaire, elle ne fournit pas des informations satisfaisantes sur les étapes qui aboutissent à des outils évaluatifs valides. C'est la carence de cet aspect qui nous pousse à investir la validation de De Ketele et Gérard (2005) pour valider les situations évaluatives. Néanmoins, la validité des outils de mesure et de jugement requiert l'exclusion de plusieurs biais qui faussent tout le processus évaluatif. Bien qu'ils soient évoqués brièvement dans certains éléments de ce chapitre, nous les présentons de manière plus détaillée dans ce qui suit.

#### 4. Les biais de l'évaluation sommative

L'évaluation sommative et les résultats auxquels elle aboutit pourraient être biaisés par plusieurs facteurs. Par *biais*, nous signifions les éléments qui affectent le processus évaluatif et en altèrent les résultats. En vue de les repérer, nous nous référons au travail de Paquette-Côté et Raïche (2013) qui insistent sur l'exclusion de certains biais pour assurer la validité de l'inférence de la notation. Selon ces auteurs, ces biais pourraient être liés à l'évaluateur, aux évalués ou à certaines conditions environnementales. Jorro (2000) qui s'intéresse aux *failles* de l'évaluation sommative dans son ouvrage *L'enseignant et l'évaluation- Des gestes évaluatifs en question* aborde également les *biais* ou *les effets* qui nuisent aux pratiques de correction et de notation, des effets que des auteurs tels qu'Abernot (1988), Bonniol (1981), Cardinet (1992), Hadji (1992), Landelle (1988), Noizet et Caverni (1978), Paristot (1988)<sup>54</sup>... ont antérieurement abordés. Quant à Tagliante (2005), elle les désigne par les *parasites*<sup>55</sup> qui affectent la fidélité de la notation. En nous inspirant de ces auteurs cités supra, nous tentons, à travers les éléments qui suivent, d'évoquer certains biais qui pourraient altérer la qualité de l'évaluation, en les intégrant dans des recommandations classées en fonction des divers moments du processus évaluatif.

##### 4.1. La planification

Au cours de cette première phase, des précisions quant au déroulement de l'évaluation sont nécessaires. En effet, certains éléments -s'ils ne sont pas déterminés préalablement- présentent des risques majeurs qui atteignent la validité du processus évaluatif. Pour pallier ces biais, il faut :

---

<sup>54</sup> Tous ces auteurs sont cités par Jorro (2000).

<sup>55</sup> L'auteur qualifie les effets qui nuisent à la qualité de la notation par *les critères parasites de la fidélité*.

- Unifier les épreuves évaluant les apprenants qui suivent le même programme : Paquette-Côté et Raïche (2013) trouvent nécessaire de proposer aux apprenants ayant le même programme d'études les mêmes activités évaluatives, même si la matière en question est assurée par des enseignants distincts. Cette indication semble incontestable si nous ne négligeons pas les objectifs visés par ces enseignants qui sont communs, ainsi que les exigences et les seuils de réussite. En effet, les enseignants -même en empruntant des voies différentes- sont censés atteindre les mêmes objectifs. En évaluant les apprenants différemment sur des apprentissages identiques, nous n'obtenons pas des résultats valides étant donné que ceux-ci révéleraient les aptitudes des apprenants du même niveau scolaire dans des situations différentes. C'est à ce niveau que la concertation des enseignants chargés des groupes du même niveau devient indispensable pour un consensus sur les épreuves, les exigences, les seuils de réussite... Mentionnons que cette coordination assure plus de fidélité aux objectifs du programme, ce qui ne serait pas le cas dans les situations d'évaluations distinctes selon les enseignants qui affectent considérablement leurs pratiques, les poussant même à économiser les contenus et l'énergie, étant donné qu'ils sont les évaluateurs de leurs propres enseignements.
- Vérifier le planning de passation : les apprenants dans l'univers scolaire n'étudient pas une seule matière concernée par l'évaluation. Ils sont alors évalués dans plusieurs disciplines suivant un programme précis. Les enseignants de la matière doivent s'assurer du nombre d'examens prévus par jour, et de l'heure de passation de l'épreuve de la matière dont ils sont chargés. En effet, ces éléments jouent un rôle primordial vu qu'un grand nombre d'épreuves prévues le même jour déstabilise la concentration des apprenants qui sont contraints à privilégier une discipline au détriment d'une autre.
- Préciser les normes de réalisation et de présentation : toutes les instructions liées à la réalisation et à la présentation des travaux doivent être précisées pour une présentation identique des travaux. Ces instructions et consignes garantissent une présentation qui ne favorise pas certaines productions étant donné que la qualité de la présentation est un facteur qui surestime certaines productions et sous-évalue d'autres.
- Étudier le moment et la durée de l'épreuve : il est important que les enseignants discutent préalablement pour prendre des décisions relatives à la durée et au moment de l'épreuve. S'agissant de la durée, il est important que les évaluateurs choisissent une durée appropriée qui permet aux apprenants de répondre aux questions prévues, et ce,

en tenant compte du nombre de questions proposé et du niveau de difficulté des situations sélectionnées. Souvent, ce sont les élèves moyens qui servent de référence pour déterminer la durée à impartir à une épreuve : les élèves excellents consomment peu de temps pour résoudre les activités proposées alors que les faibles mettent plus que le temps nécessaire pour exécuter la même tâche. Pour ce qui est du moment, il est connu que certains moments sont plus propices à un meilleur rendement scolaire, d'où la nécessité de choisir un moment opportun à une meilleure évaluation. Ce moment prend en considération la nature des activités que les épreuves évaluatives comportent : les épreuves qui mobilisent des opérations cognitives de haut niveau –contrairement à celles réduites à la simple restitution des connaissances- ne pourraient pas être programmées à n'importe quel moment du jour.

- S'occuper des conditions matérielles de la passation : certaines conditions matérielles du déroulement, telles que l'aménagement des salles, la luminosité, la température..., sont à prendre en considération dès la phase de planification.

Ces éléments que nous venons de citer permettent d'éviter certains biais nuisibles à la validité du processus évaluatif. Entre autres, citons : une évaluation subjective des apprenants du même établissement étant donné que des épreuves différentes signifient des niveaux de difficulté variée, un nombre élevé d'évaluations pendant le même jour étant donné que la première épreuve épuise l'énergie des apprenants altérant ainsi leur rendement dans les épreuves qui suivent, une présentation variée des productions qui privilégie les apprenants de certaines catégories ...

#### **4.2. La conception des épreuves**

Dans cette deuxième phase, les enseignants doivent concevoir des outils évaluatifs (des items ou des activités) évaluant les compétences des apprenants. Pour éviter les éléments qui biaisent l'évaluation, il est nécessaire de :

- Préciser les consignes : lors de l'élaboration des questions, les enseignants doivent nécessairement formuler leurs questions de manière claire, précise et pertinente. La précision des questions est garante de réponses répondant aux attentes des évaluateurs. A travers chacune des questions posées, les enseignants veulent évaluer des connaissances et des compétences précises. En posant les questions de manière erronée, les enseignants obtiennent des réponses autres que celles auxquelles ils s'attendent, ce qui biaise pleinement leur évaluation. En effet, les réponses incorrectes des apprenants

dans ce cas-ci ne seraient pas nécessairement dues à leur incompetence dans la mesure où ils pourraient formuler de bonnes réponses si la question était posée autrement. C'est pour cette raison que la reformulation des questions et leur explication aux apprenants même le jour de la passation ne sont pas objectées par tous les enseignants, qui se rendent tard de certaines erreurs de formulation<sup>56</sup>.

- Préétablir les normes de notation : lors de l'élaboration des épreuves, les enseignants ont à élaborer les barèmes de notation ou les grilles de correction, et ce, en fonction de la démarche adoptée. S'agissant de ces outils, leur précision est garante d'une évaluation équitable des productions des apprenants ayant passé les épreuves conçues. Il serait aberrant que les enseignants répartissent différemment les notes sur les diverses questions dans la mesure où certains pourraient se référer à des critères autres que l'importance des questions. Si les notes réparties sur les items proposés reflètent souvent leur importance, certains enseignants préfèrent concevoir leurs barèmes en fonction des réponses de leurs apprenants. La question à laquelle un nombre élevé, ou un groupe privilégié, répond incorrectement est sous-évaluée si les enseignants veulent aider leurs apprenants, mais surévaluée s'ils veulent les sanctionner. Il en est de même pour les grilles de correction dont la précision est garante d'une évaluation assez objective, étant donné que sans ces grilles les enseignants se livrent à des appréciations subjectives et personnelles sur les productions des apprenants, et donc à une évaluation subjective. Il est nécessaire également que les enseignants se chargent d'appliquer soigneusement les barèmes et grilles négociées avec les enseignants assurant la même matière, notamment que cette négociation ne doit pas négliger toutes les réponses éventuelles des apprenants.

Ces quelques éléments que nous avons cités garantissent une évaluation objective des connaissances et compétences réelles des apprenants. En les appliquant, nous évitons deux biais essentiels qui nuisent à la validité de l'évaluation : la subjectivité des outils de notation, barèmes et grilles, ainsi que l'inadéquation des réponses à ce que les enseignants veulent évaluer. Après avoir précisé les biais qui peuvent être évités lors de la planification et de la conception des épreuves, nous découvrons ci-dessous ceux à éviter au moment de la correction.

---

<sup>56</sup> Cette remarque n'est pas valable pour toutes les disciplines. Ainsi, en évaluation des langues étrangères, la compréhension de la consigne, qui est nécessaire pour répondre aux questions posées, est un élément évalué par certains enseignants. Ceux-ci doivent être très attentifs lors de la reformulation de la consigne ou son explication pour ne pas orienter les apprenants vers la bonne réponse.

### 4.3. La correction des copies

La correction est le moment le plus important car selon Gérard, c'est le moment qui « doit permettre de donner sens au travail de l'élève » (2009, p. 121). D'ailleurs, les failles que les auteurs (Tagliante, 2005 ; Jorro, 2000...) énumèrent sont liées à ce moment. A ce niveau de notre étude, nous intégrons les effets qui affectent la fidélité des notes, et par-là leur validité. Les enseignants durant cette phase doivent :

- Appliquer les barèmes et les grilles préétablis : lorsque la correction des copies est assurée par plusieurs enseignants, ceux-ci doivent appliquer le plus fidèlement possible les barèmes de notation ou les grilles de correction préétablis, dans la mesure où tout changement, aussi minime semble-t-il, biaise considérablement la correction. Il en est de même si un seul enseignant corrige les copies étant donné que tout arrangement altère la qualité de la correction.
- Éviter une grande sévérité : les enseignants tentent de corriger les copies sans adopter un style très sévère étant donné qu'une grande sévérité pourrait accorder à une production moins qu'elle mérite. Ceci ne signifie aucunement une tolérance exagérée accordant aux apprenants plus qu'ils ne méritent.
- Éviter les arrangements et les tendances répandues : il est nécessaire que les enseignants évaluent les travaux de leurs apprenants indépendamment de toutes considérations personnelles, professionnelles, sociales... A ce niveau, certains enseignants attribuent leurs notes de façon à obtenir une répartition gaussienne<sup>57</sup>, avec une majorité moyenne, quelques apprenants excellents et quelques apprenants faibles. Une telle répartition est, aux yeux de certains, le prix à payer pour préserver sa réputation d'enseignant modèle. D'autres enseignants, pour préserver leur réputation, attribuent à leurs apprenants de bonnes notes pour paraître de bons enseignants, aux yeux des apprenants et de leurs parents.

Ces éléments peuvent écarter quelques biais qui affectent la notation et la correction des copies. Dans ce qui suit, nous les citons en nous inspirant de Tagliante (2005), qui les explique brièvement :

---

<sup>57</sup> La courbe de Gauss, ou la courbe en cloche, correspond à la forme de la courbe représentant graphiquement les données. Appliquée aux notes obtenues par les apprenants dans les épreuves d'évaluation, elle reflète une minorité d'apprenants faibles, une majorité de moyens et une minorité d'excellents.

- Effet de fatigue : cet effet est lié à l'état physiologique et psychologique de l'enseignant qui a des effets sur les notes attribuées aux apprenants. S'agissant de l'état physiologique, il est connu que les enseignants corrigent mieux les copies de leurs apprenants le matin (Tagliante, 2005), contrairement aux autres moments de la journée où la fatigue altère la qualité de leur correction. Il n'est pas question que de la fatigue physiologique étant donné que les émotions négatives des enseignants impactent indéniablement sur la correction : une humeur exacerbée pendant la correction des copies dévalorise sans doute les productions des évalués. Les enseignants doivent alors choisir des moments appropriés pour une appréciation assez objective des réponses des apprenants.
- Effet de contraste : cet effet résulte de la confrontation que les enseignants effectuent même inconsciemment entre les productions corrigées. Ainsi, toute production corrigée après une production excellente souffre d'une appréciation dévalorisante. L'inverse est tout-à-fait correct vu qu'une production moyenne serait valorisée si la production qui la précède s'avère très faible.
- Effet de l'ordre : cette fois-ci, les biais ne sont pas dus aux productions précédant ou suivant la production corrigée, mais à l'ordre de la correction des copies. Il est connu que les enseignants ont tendance à devenir de plus en plus sévères vers la fin de la correction, ce qui signifie que les copies corrigées au début souffrent moins de la sévérité des enseignants.
- Effet de contamination : dans de nombreux cas, les enseignants se laissent influencés par le dossier scolaire de leurs apprenants ou par leurs notes antérieures (Jorro, 2000). De bons dossiers scolaires avantagent largement les apprenants dont les erreurs ne seraient plus perçues du même œil que celles que commettent des apprenants de faible niveau. S'agissant des notes antérieures des apprenants, contrairement à Jorro (2000) qui leur associe un effet de contamination étant donné qu'ils impactent sur la correction des copies, Tagliante (2005) les aborde au niveau de l'effet de stéréotype.
- Effet de stéréotype : au niveau de cet effet, les enseignants après une ou des évaluations antérieurement introduites ont tendance à catégoriser les apprenants ou à les classer dans des catégories particulières. Un apprenant qui est jugé excellent dans certaines épreuves bénéficie considérablement d'une correction moins sévère que ses pairs classés dans les catégories des incompetents. Cet effet présente un risque majeur car il

freine les apprenants qui s'améliorent et progressent en les emprisonnant dans une case dont ils ne peuvent s'échapper aisément.

- Effet de halo : tel que Tagliante (2005) le présente, cet effet est lié à la présentation de la copie ou du travail des apprenants. En effet, les enseignants ont tendance à mieux noter un travail soigné avec une écriture lisible et claire, contrairement à un travail mal présenté qui, en dépit de la qualité des informations fournies, est souvent sous-apprécié. A cet élément, Jorro (2000) ajoute d'autres facteurs, comme l'aspect de l'apprenant, qui pourraient impacter une évaluation objective des compétences des évalués. L'influence de ce facteur (aspect du candidat) sur la correction et la notation est incontestable. Personne ne réfute catégoriquement le rôle que jouent le charisme, la tenue, les attitudes, les manières... des apprenants dans l'appréciation de leurs travaux lorsqu'il est question d'une évaluation orale. Les enseignants se laissent, même inconsciemment, impressionnés par ces facteurs qui empêchent une évaluation authentique des compétences des apprenants étant donné que ces biais parasitent le jugement évaluatif.
- Effet goutte d'eau : cet effet est lié à des erreurs qui deviennent inadmissibles pour les enseignants car elles sont précédées par un nombre élevé d'incorrections. Dans ce cas-ci, l'erreur n'est pas jugée objectivement par les correcteurs qui ne s'intéressent plus à sa nature ni à la possibilité de la croiser dans une production d'un apprenant d'un niveau donné, mais c'est sa position dans la production qui la rend intolérable, étant donné que cette même erreur serait traitée différemment si elle figurait au début de la production. La correction de toute la production se trouve influencée par cette erreur qui s'attire une plus grande sévérité de la part des enseignants, ce qui ne convient pas dans une évaluation réelle des apprentissages scolaires.
- L'évaluation externe : ce dernier élément que nous citons se rapporte aux comportements adoptés par certains enseignants qui, soucieux de leur réputation, accordent aux apprenants des notes surestiment ou sous-évaluent leurs travaux. Dans le premier cas, les enseignants veulent paraître compétents car souvent ils sont évalués à travers les résultats de leurs apprenants. En leur accordant de bons résultats, ils contribuent à hausser leur valeur et leur réputation. Dans le cas échéant, les correcteurs attribuent des notes qui sous-estiment les productions corrigées pour ne pas paraître laxistes. Dans les deux cas, la correction des enseignants est altérée et leur notation est loin de refléter le travail réel des apprenants.

Tous ces effets que nous venons de citer affectent la fidélité de la notation et par-là sa validité étant donné que les notes accordées par les correcteurs ne traduisent pas la compétence évaluée. C'est pour les éviter que certains auteurs tels que Gérard (2009) proposent des stratégies permettant aux enseignants de corriger les productions des apprenants sans être perturbés par les effets cités. Contrairement aux recommandations que nous avons énumérées pour aborder certains biais qui altèrent le processus évaluatif, les stratégies préconisées par Gérard (2009) peuvent être appliquées au moment de la correction. Il s'agit de procéder à une correction fragmentée par critère : la maîtrise de chaque critère est vérifiée dans toutes les productions avant de passer au critère suivant. Cette stratégie qui est applicable avec les grilles de correction n'est pas la seule à laquelle les enseignants peuvent recourir. Nombreux sont les correcteurs qui procèdent à une correction anonyme des copies pour ne pas être influencés par la connaissance antérieure des apprenants ou de leurs compétences qui biaise la correction et la notation. Cette stratégie est certainement généralisée étant donné que dans la correction des concours, des tests et des épreuves officielles et professionnelles, la correction anonyme des copies est souvent appliquée. A ce mode de correction, s'ajoute une deuxième démarche soutenant encore la validité de la correction et par-là celle des notes : la multiplication des correcteurs : une copie est corrigée par deux correcteurs, le troisième n'étant sollicité qu'en cas d'écart flagrant entre les deux notes attribuées.

Au sein des établissements scolaires, ces démarches peuvent être appliquées par les enseignants qui quêtent une évaluation assez correcte des compétences des apprenants. Ceux-ci organisent des séances de coordination pour élaborer les épreuves évaluatives en tenant compte des objectifs institutionnels et des contenus traités, concevoir les barèmes ou les grilles de correction à appliquer, étudier les conditions de passation (moment, durée, salle...), et se concerter sur les modalités de correction (correction de chaque production par un ou plusieurs correcteurs).

En somme, les biais affectant la validité de l'évaluation sommative que nous venons de citer ne sont pas exhaustifs étant donné que les ouvrages consultés n'en fournissent pas une liste détaillée. Souvent, en évoquant les failles de cette évaluation, la littérature scientifique aborde les effets qui altèrent la qualité des notes attribuées par les enseignants, les réduisant ainsi au moment de la correction. En abordant les autres éléments, nous voulons signaler que les éléments qui biaisent le processus évaluatif excèdent le moment de la correction car ils sont intrinsèques à toutes les phases du processus. Toute faille affectant l'opération évaluative dès son commencement altère la qualité de l'évaluation et celle des notes attribuées aux apprenants.



Bien qu'une évaluation sommative exempte de biais semble inatteignable, les enseignants doivent prendre toutes les mesures nécessaires pour juger le plus fidèlement les productions évaluées.

### **Conclusion**

De tout ce qui a été évoqué dans les éléments précédents, il s'est avéré que la validité n'est pas une notion aisée à définir, d'où le désaccord entre les auteurs sur la conception à adopter. En effet, deux conceptions définitives sont repérées : la première jugée classique associe la validité à une qualité du test, plus précisément à une caractéristique des instruments évaluatifs intégrés en évaluation des apprentissages (Noël-Jothy & Sampsonis, 2006 ; Tagliante, 2005 ; Riba & Mègre, 2015...), alors que la seconde affirme que la validité est liée notamment à la qualité de l'interprétation des résultats (Raïche, Ndinga & Meunier, 2013) en vue d'un jugement valide sur les apprentissages évalués.

Chaque conception a ses partisans dont certains rejettent la conception différente, créant par-là une certaine opposition entre les deux. Ainsi, Raïche, Ndinga et Meunier (2013) objectent la conception classique qualifiée de fautive tandis que Borsboom, Mellenbergh et Van Heerden (2004) réfutent la conception nouvelle, plaidant ainsi pour le retour à la conception classique de la validité. Dans notre étude, nous préférons à cette opposition inutile une complémentarité entre les deux positions étant donné qu'elles ne s'excluent pas. En effet, la validité de l'interprétation débute là où s'achève celle de la mesure. La notion est alors prise pour signifier la capacité de tous les instruments évaluatifs intégrés tout au long du processus évaluatif en vue d'aboutir à des résultats valides, reflétant le plus fidèlement possible les compétences des apprenants évalués.

Cette tentative de cerner les frontières de la notion clé de notre étude est complétée par une explication des différents types que la validité englobe. Ces types sont si nombreux qu'aucun auteur n'en présente une liste exhaustive : les travaux se contentent d'en traiter les plus connus. A l'instar des travaux consultés et cités dans notre étude (De Ketele, 2010 ; Demeuse & Henry, 2004...), nous nous intéressons aux types qui interpellent l'attention d'un nombre considérable d'auteurs. Chacun de ces types se distingue par une dimension particulière : la validité prédictive est liée à la capacité du test à prédire les résultats des apprenants dans des tests ultérieurs, celle du contenu correspond au contenu du test et à la

représentativité des items... De tous les types cités, la validité du contenu est la plus liée à la validité dont il est couramment question dans les travaux scientifiques. D'ailleurs, la conception classique qui associe la validité à une qualité du test qui mesure ce qui doit être mesuré présente des similitudes frappantes avec la validité du contenu.

La présentation des divers types cités et de leurs particularités est nécessaire car, d'une part, elle contribue à mieux éclaircir la notion de validité et, d'autre part, elle introduit les procédures de validation appuyées sur ces types. En effet, à travers ce chapitre, nous avons traité, outre la définition de la validité et ses divers types, les procédures de validation, c'est-à-dire, qui prouvent la validité des épreuves conçues par les enseignants. Ces procédures de validation, nous les avons abordées pour en repérer les phases que la conception des dispositifs évaluatifs rigoureux doit suivre, valider et expérimenter les épreuves évaluatives conçues et identifier les biais qui faussent toute évaluation valide. De toutes les validations évoquées, celle de Kane (2006) a été longuement expliquée. Investie dans de nombreuses études, déjà citées dans le corps de notre chapitre, nous en servirons pour concevoir certains critères des grilles que nous exploiterons dans notre partie pratique en vue de l'analyse des épreuves écrites proposées en évaluation de FLE, au cycle moyen.

Bien que nous reconnaissons l'utilité de la validation de Kane (2006), nous exploiterons notamment celle de De Ketele et Gérard (2005) qui valide les épreuves écrites conçues dans le cadre de l'APC, actuellement appliquée dans notre système éducatif. En effet, nous nous en inspirerons pour concevoir des outils évaluatifs reconnus valides par les auteurs en veillant à répondre à toutes les exigences de cette approche en matière d'évaluation scolaire. Cependant, avant de passer à la validation des épreuves constituant notre corpus et à celle des épreuves que nous proposerons dans notre partie pratique, nous jugeons nécessaire de consacrer le dernier chapitre de notre volet théorique aux étapes de la conception des outils évaluatifs (mesure et jugement) qui répondent au contexte actuel. C'est ce dernier chapitre qui trace les pistes des dispositifs que nous proposerons pour une évaluation des compétences des apprenants algériens en FLE.

**CHAPITRE 3 : DES ETAPES POUR LA  
CONCEPTION DES DISPOSITIFS  
EVALUATIFS**

## Introduction

Dans les deux chapitres précédents, nous avons précisé que l'évaluation sommative -qui est distincte des formes formative et certificative- consiste en des épreuves orales et écrites introduites au terme d'une série d'apprentissages pour en vérifier la maîtrise. Dans un souci de validité, qui requiert essentiellement la représentativité des items retenus, les enseignants s'efforcent à sélectionner des activités conformes aux exigences des pédagogies en vigueur. Tout décalage entre les objectifs institutionnels et les objectifs évalués, ou entre les pratiques enseignantes et les pratiques évaluatives est à éviter pour une évaluation valide et rigoureuse, répondant aux attentes des acteurs impliqués (apprenants, parents, enseignants, recruteurs...). Etant donné que chaque pédagogie ou approche interpelle des modalités évaluatives particulières, il devient nécessaire de déterminer quelles activités évaluatives nous devons retenir dans le cadre de l'APC, actuellement appliquée à l'enseignement/apprentissage du FLE en Algérie. C'est en déterminant les activités requises dans cette approche que nous arrivons à proposer un dispositif valide évaluant réellement les compétences des apprenants du cycle moyen en FLE.

En vue d'atteindre cet objectif, nous aborderons dans ce troisième chapitre les étapes à suivre pour concevoir un dispositif évaluatif conforme aux exigences de l'APC, approche préconisée. Avant de les détailler, nous débutons par une définition d'un concept clé dans notre étude (dispositif), en nous appuyant sur la thèse de Roy (2017) intitulée « *Les dispositifs d'évaluation certificative mis en œuvre par les enseignants pour évaluer les attitudes dans un programme d'études collégiales : étude de cas* ». Comme cet auteur, outre la définition du concept dispositif, retrace dans le même travail toutes les étapes que les évaluateurs à suivre en vue de la conception de dispositifs évaluatifs, nous nous en inspirons pour préciser les moments à respecter pour obtenir des outils évaluatifs adaptés à ce qui est préconisé dans les programmes institutionnels. Au sein de nos explications de ces étapes, nous infiltrons des réponses à toutes les questions auxquelles il est indispensable de répondre dans notre étude.

Les premières questions ont trait aux éléments évaluables, à évaluer et réellement évalués dans les programmes actuels. Concernant les éléments évaluables, nous tenterons de déterminer tous les objets dont une évaluation est possible dans le cadre de l'APC. Si nous estimons qu'une évaluation des connaissances linguistiques est inadmissible dans une perspective de compétences, il devient important de préciser les objets à évaluer : faut-il évaluer les connaissances ou les compétences ? Est-il possible d'intégrer les deux composantes dans une épreuve ? De cette question, jaillit une autre aussi importante : est-il possible d'évaluer les

compétences ? Les éléments évaluables et ceux à évaluer doivent être préalablement précisés pour dessiner la structure des épreuves évaluatives adaptées à l'approche appliquée. Quant aux éléments réellement évalués, nous les évoquerons brièvement dans ce chapitre car nous les étudierons, dans notre partie pratique, à travers l'analyse des épreuves proposées au cycle moyen pour évaluer les apprenants en FLE.

S'agissant des deuxièmes questions, elles concernent les outils évaluatifs que l'APC exige. Les objets à évaluer selon cette approche étant précisés, nous mentionnons comment les évaluer. Nous précisons ainsi les outils auxquels les évaluateurs doivent recourir pour évaluer le plus fidèlement possible les connaissances ou les compétences évaluées. Généralement, les situations sont les instruments de mesure que les évaluateurs jugent adaptées à l'APC car elles aboutissent à des productions concrètes révélant les profils des évalués. Néanmoins, le milieu scolaire continue à valider les apprentissages des apprenants sur la base d'épreuves constituées d'exercices structuraux évaluant leur compétence linguistique. De ces outils cités, nous tentons de retenir ceux qui semblent plus adéquats aux attentes des évaluateurs.

Soulignons que la précision des outils évaluatifs interpelle celle des outils correctifs. En effet, la tâche des évaluateurs ne se réduit pas à proposer aux apprenants des activités pour les évaluer, car elle comprend aussi la correction de leurs travaux ou productions pour émettre des jugements fondés sur leurs acquis. Nous tenterons alors de recenser les instruments auxquels les évaluateurs doivent recourir pour évaluer les réponses de leurs apprenants dans les épreuves évaluatives proposées. Quels outils devons-nous investir lors de la correction des copies des apprenants ? Quelles stratégies écartent-elles les biais qui faussent les notes des évalués ?

Ces questions non-exhaustives nous semblent importantes pour aboutir à une évaluation authentique<sup>58</sup> des apprentissages programmés, à travers des outils appropriés. Notre objectif, certes ambitieux, vaut la peine d'être poursuivi notamment que la littérature scientifique récente ne cesse de dénoncer le décalage entre les pratiques exigées et les pratiques effectives des enseignants en matière d'évaluation scolaire. Cet écart constaté ne pourrait être comblé qu'à travers une étude expliquant les modalités requises pour une évaluation conforme aux instructions de l'APC, actuellement appliquée dans de nombreux systèmes éducatifs.

### **1. Qu'est-ce qu'un dispositif d'évaluation ?**

---

<sup>58</sup> Une évaluation *authentique* dans le sens propre du terme n'existe pas. Par ce qualificatif, nous entendons une évaluation valide qui donne un aperçu assez fidèle aux compétences des évalués, à un moment donné de leur apprentissage.

Roy définit le dispositif d'évaluation comme un « *ensemble organisé des éléments rattachés à la démarche d'évaluation permettant de certifier une compétence au terme d'un cours* » (2017, p. 9). Ces éléments sont tous les outils qui mesurent les apprentissages (questions et activités proposées), décrivent les résultats obtenus (notes, annotations...) et les communiquent aux acteurs sociaux (bulletins...). Les enseignants ou examinateurs les intègrent tout au long du processus évaluatif qui s'étale sur les phases suivantes : l'intention, la mesure, le jugement et la décision. Ces quatre phases constitutives du processus de l'évaluation, que nous devons à Cardinet (1975), sont reprises par des auteurs tels que Cuq (2003), Bélair (2014) ainsi que Leroux et Bélair (2015), qui ont conçu un schéma pour résumer les étapes citées supra telles que Cardinet les cite. En quoi consiste chacune de ces étapes ?

L'intention correspond à la *planification*, un concept que Durand et Chouinard (2006) préfèrent à celui d'intention. Elle englobe des précisions sur les objets à évaluer, les objectifs de l'évaluation, les moments de son intégration, les activités appropriées... La deuxième étape du processus la mesure, désignée par *prise d'information* par Lasnier (2014) inclut la conception des activités évaluatives, des outils de jugement... Elle permet ainsi de concevoir et de mettre en œuvre ces outils en vue d'obtenir des productions ou des preuves ce qui est évalué. C'est la phase de jugement qui assure la correction et l'interprétation des preuves collectées à travers les outils antérieurement conçus. Elle permet ainsi de fournir des jugements sur les connaissances et les compétences évaluées pour préparer la prise de décision. Les décisions que les évaluateurs prennent concernent les notes à attribuer, les ajustements à prévoir...

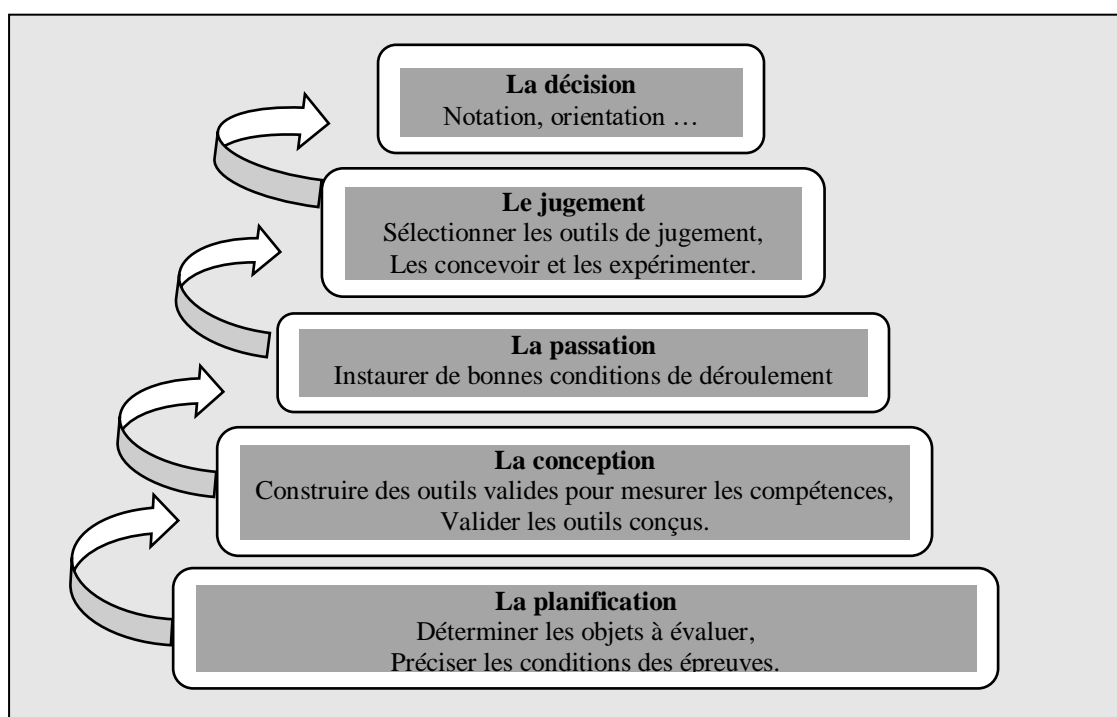
Toutes ces étapes, que nous ne traitons pas abondamment à ce niveau, sont également citées par Roy (2017) pour évoquer les moments de la conception des dispositifs de l'évaluation certificative, que l'auteur ne distingue pas de la forme sommative. Ce travail de Roy (2017), nous nous y référons dans notre étude en le soutenant par les étapes de conception des tests et examens telles que le « *Manuel pour l'élaboration et la passation de tests et d'examens de langue* » les précise. En effet, ce document du Conseil de l'Europe aborde l'élaboration des outils de mesure, ne constituant qu'un moment du processus évaluatif. Ce moment-ci -qui englobe la planification, la conception, l'expérimentation et l'information des partenaires<sup>59</sup>- ne

---

<sup>59</sup> La planification est le moment de répondre à des questions décisives portant sur les candidats, les objectifs du test, les objets à évaluer, les seuils de réussite... Pour la conception, elle se rapporte à l'élaboration et la sélection des épreuves que le test doit inclure en tenant compte des critères professionnels de toute évaluation. Quant à l'expérimentation, elle vise la vérification de la pertinence des instruments de mesure et de correction. Enfin, l'information des acteurs impliqués consiste à fournir des versions aux enseignants et à préparer les apprenants.

concerne que la conception des activités évaluatives liées à la *mesure*, telle que Roy (2017) la conçoit. Il est précédé par la *décision de concevoir* le test et suivi par la *passation*, la *correction*, la *notation* et la *délivrance des résultats*.

Certaines étapes sont similaires dans les deux travaux en question telles que la correction qui peut être associée au jugement, et la notation qui rejoint la décision. Il existe cependant une étape qui ne figure pas dans le processus de Roy (2017) et que le Conseil de l'Europe (2011) aborde : la passation. C'est en fusionnant les deux perspectives que nous retenons les phases de la conception de dispositif évaluant les compétences des collégiens algériens en FLE. Nous les représentons dans la figure qui suit :



**Figure 2: Schéma des étapes de la conception de dispositifs valides**

A travers la précision des étapes constitutives du processus évaluatif, notre intention est de déterminer dans la suite de notre travail les outils relatifs aux moments de la démarche évaluative, étant donné que ce sont ces outils qui constituent les dispositifs d'évaluation. Quels sont donc ces outils ? Lesquels sont plus adaptés à l'APC, actuellement en vigueur ?

## **2. La planification : une étape-clé vers un dispositif valide**

Cette première phase fonde les piliers d'un dispositif d'évaluation valide, et ce en répondant à une série de questions déterminant les objectifs de l'évaluation menée, les apprentissages à évaluer, les moments de son intégration... Ce sont les programmes et les

documents officiels (guides) explicitant les modalités d'enseignement et d'évaluation de chaque discipline qui constituent les références à consulter par les enseignants en vue de parvenir à des réponses satisfaisantes. De tous les éléments que cette première phase ne doit pas négliger, les objectifs visés par l'enseignement/apprentissage déterminent en grande partie la nature des outils évaluatifs (items, exercices d'application ou situations) à concevoir par les enseignants-évaluateurs. Si le contexte éducatif actuel connaît une rupture avec la pédagogie par objectifs (PPO) qui privilégie des modes évaluatifs particuliers, il devient nécessaire de consacrer un volet dans le cadre de notre étude à l'évaluation dans le cadre de l'APC dont les pratiques évaluatives diffèrent de celles de l'ancienne pédagogie. Que faut-il évaluer dans le cadre de cette approche ? Et par quels moyens ?

### **2.1. Que faut-il évaluer dans le cadre de l'APC ?**

Selon De Ketele et Dufays (2003), les approches et les pédagogies appliquées conditionnent la sélection des pratiques évaluatives : chaque pédagogie entraîne des instruments évaluatifs précis. Ainsi, dans le cadre de la PPO, les évaluateurs sélectionnent un échantillon représentatif des objectifs spécifiques et opérationnels -visés par l'enseignement/apprentissage- qui sont traduits en question. Cependant, dans le cadre de l'APC, qui développe des compétences chez les apprenants, ce sont ces compétences que les enseignants évaluent pour en mesurer la maîtrise. Que ce concept compétence désigne-t-il en évaluation de langues ?

#### **2.1.1. La compétence : des définitions**

La définition de la compétence, un concept polysémique<sup>60</sup> qui n'est pas très récent dans le monde pédagogique, « *est loin d'être, à ce jour, tout-à-fait stabilisée* » (Delory, 2002, p. 23), d'où la nécessité de la préciser en nous référant à la littérature scientifique. En vue de définir cette notion dans le milieu éducatif, Gérard (2009) s'inspire des trois dimensions qui fondent sa signification dans le milieu professionnel : l'originalité, l'efficacité et l'intégration. En effet, la compétence des évalués se manifeste dans des réponses *originales* et spécifiques à des situations -issues de la même famille- qui doivent être inédites ; *efficaces*, étant donné que les réponses doivent être exactes et utiles ; et *intégrées*, dans la mesure où les apprenants sont appelés à mobiliser leurs acquis pour faire face à ces situations. Ces trois dimensions contribuent à la définition de la compétence qui consiste à « *mobiliser un ensemble intégré de*

---

<sup>60</sup> Delory (1991) est l'un des auteurs ayant évoqué la polysémie du mot qui n'est pas encore clairement défini.



*ressources pour résoudre des situations-problèmes* » (Gérard, 2009, p. 10). Cette première définition qui insiste sur deux dimensions de la compétence (mobilisation et efficacité) s'approche de celle de la Communauté française de Belgique citée par Delory. La compétence est définie comme une « *aptitude à mettre en œuvre un ensemble organisé de savoirs, des savoir-faire d'aptitudes permettant d'accomplir un certain nombre de tâches* » (Communauté française de Belgique, 1997, Citée dans Delory, 2002, p. 24). Delory, qui précise les ressources à mobiliser par les apprenants, rejoint Gérard (2009) dans sa conception de la compétence comme une capacité assurant le transfert de ce qui est appris dans des situations nouvelles.

Bien que la compétence soit souvent opposée à la connaissance, des auteurs tels que Perrenoud (1999), Hadji (2015) et Médioni (2016) ne s'attardent pas sur la distinction opérée entre connaissances et compétences vu que le besoin de former des apprenants compétents s'impose ces dernières années valorisant ainsi les compétences dans le milieu scolaire. Les enseignants ne visent plus la transmission de connaissances que les apprenants assimilent puis restituent : leur nouvel objectif est de développer la capacité de mettre en œuvre ces connaissances, savoir-faire et stratégies pour résoudre diverses situations-problèmes. Dépasser cette opposition ne réfute pas le lien qui existe entre les deux notions dans la mesure où les connaissances constituent une condition indispensable aux compétences. Cette importance des connaissances ou ressources<sup>61</sup> est mise en valeur par la définition de Tardif qui décrit la compétence comme « *un savoir-agir complexe prenant appui sur la mobilisation et la combinaison efficaces d'une variété de ressources* » (2006, p. 22). Or, il est connu que ces connaissances –même si elles sont procédurales- ne suffisent pas face à la complexité des situations, il devient légitime de mettre l'accent sur les compétences dans les pratiques pédagogiques et évaluatives des enseignants.

Selon Hadji (2015), trois idées sont intégrées pour définir les compétences dans les milieux éducatifs : la situation, la complexité et la performance. En nous inspirant de son ouvrage « *L'évaluation à l'école. Pour la réussite de tous les élèves* », nous explicitons ces trois idées dans ce qui suit. La situation, première idée listée par Hadji, est un élément-clé dans la définition de la compétence qui est contextualisée ; autrement dit, la compétence se concrétise dans un contexte ou un environnement particulier. En effet, aucune définition formulée de cette notion de compétence ne peut écarter la situation. Scallon, qui conçoit la compétence comme

---

<sup>61</sup> La notion de ressources telle que Gérard (2009) l'évoque renvoie dans l'univers scolaires aux savoirs et savoir-faire. Ces ressources se répartissent en ressources internes qui renvoient aux acquis des apprenants (savoirs et savoir-faire) et en ressources externes liées aux ressources matérielles, sociales et procédurales (Gérard, 2009).

un « *savoir-agir en situation* »<sup>62</sup> estime qu'« *on parle de compétence lorsqu'un individu mobilise, c'est-à-dire utilise à bon escient, ses savoirs et ses savoir-faire dans des situations variées* » appartenant à une même famille (2007, p. 11, Cité dans Hadji, 2015, p. 48). Cette conception de Scallon -adoptée par Hadji- ne diffère pas de celle de Gérard. En effet, les trois dimensions professionnelles de la compétence évoquées supra sont repérables dans la citation de Scallon : la mobilisation s'approche de l'*intégration*, l'utilisation correcte des ressources renvoie à l'*efficacité* et la variété des situations se rapporte à l'*originalité*. A ce niveau, mentionnons que Hadji (2015) distingue la notion d'intégration de celle de mobilisation. A la base, l'intégration telle que Lenoir et Sauve, qui reprennent le Conseil supérieur de l'éducation du Québec (1991), la définissent est un « *processus par lequel un élève greffe un nouveau savoir à ses savoirs antérieurs, restructure en conséquence son univers intérieur et applique à de nouvelles situations concrètes les savoirs acquis* » (1998, cité dans Delory, 2002, p. 28). Ce processus englobe alors l'assimilation des nouvelles connaissances et le réinvestissement de tout ce qui est appris, ou sa mobilisation dans des situations originales. Ainsi conçue, l'intégration devient un processus global qui inclut la mobilisation. Ce processus de mobilisation est également évoqué par Le Boterf (1994) et Perrenoud (1999), qui définissent la compétence comme « *la capacité de mobiliser toutes sortes de ressources cognitives disponibles pour résoudre une tâche* » (Dolz, 2002, p. 140). Elle ne se réduit pas à un ensemble de connaissances déclaratives et procédurales assimilées par les apprenants étant donné que ces connaissances sont renforcées par des stratégies assurant l'analyse des situations-problèmes confrontées en vue de les résoudre.

Aux dimensions abordées par Hadji (2015), Scallon (2007) et Gérard (2009), la complexité et la performance constituent un apport des définitions de la compétence dans le domaine scolaire. Pour la complexité, Scallon –cité par Hadji- affirme que les enseignants doivent proposer « *des activités complexes comportant une certaine part d'incertitude et exigeant beaucoup d'engagement personnel* » (Scallon, 2007, p. 93, Cité dans Hadji, 2015, p. 48). Ce sont les activités mobilisant les habiletés de haut niveau -en vue de résoudre un problème donné- qui suscitent l'engagement des apprenants et leur investissement dans la tâche (Howe & Ménard, 1993). Quant à la performance, elle « *renvoie à la mise en œuvre (processus) et au résultat concret, en situation de production et d'expression, de la compétence linguistique, communicative ou culturelle* » (Cuq, 2003, p. 193). Cette définition met la lumière sur deux

---

<sup>62</sup> Cette expression de Scallon (2007, p. 101) est reprise par Hadji (2015, p. 47).

traits qui assurent la distinction entre la compétence et la performance. Le premier élément repérable est la situation : la performance est conçue comme une compétence en situation. Ce premier trait ne sera pas admis dans notre étude dans la mesure où il contredit une dimension essentielle de la compétence listée par des auteurs comme Scallon (2007), Hadji (2015) et Gérard (2009). Quant au second, il concerne le caractère concret de la performance contrairement à la compétence qui n'est pas observable. Ce caractère concret est à l'origine des travaux de Perrenoud (1999), Allal (2000)... qui estiment que la compétence des apprenants ne pourrait être évaluée qu'à travers un ensemble de performances. Inclure dans la définition de la compétence une évaluation de la performance relève, selon Coulet (2016), d'une description de la compétence d'un point de vue évaluatif. Coulet (2016) range les citations de De Landsheere -qui conçoit la compétence comme une « *capacité d'accomplir une tâche de façon satisfaisante* » (1988<sup>63</sup>, Cité dans Aubret, Gilbert & Pigeyre, 1993, p. 32)- et de Michel et Ledru -qui la décrivent comme une « *capacité à résoudre des problèmes de manière efficace dans un contexte donné* » (1991<sup>64</sup>, cités dans Bellier, 1999, p. 235)- dans les définitions évaluatives qui insistent sur les jugements d'efficacité dans la réalisation de la tâche.

En somme, les auteurs ayant associé la notion de compétence à la mobilisation des ressources dans les situations appropriées présentent des repères sur son évaluation qui porte sur la capacité de mobilisation manifestée par les apprenants. De cette manière, évaluer dans le cadre scolaire ne se réduit pas à mesurer le degré de maîtrise des ressources, mais plutôt leur mobilisation dans les situations proposées. Cette évaluation devient une évaluation des processus et démarches auxquels les apprenants recourent lors de la résolution des situations-problèmes confrontées sans tenir compte du résultat ou du produit obtenu. Cette tendance s'oppose à celle des enseignants qui évaluent la compétence à travers la performance manifestée par les apprenants (des productions). Ce sont les résultats auxquels les apprenants aboutissent, ou leur rendement, qui leur permettent de juger la compétence. Ces deux tendances, non exhaustives, n'échappent pas aux critiques. Dans ce sens, nous mentionnons que des auteurs comme Kahn et Rey (2016) abordent les failles des conceptions de la compétence comme une mobilisation de ressources ou un transfert de structures.

Néanmoins, le débat scientifique ne porte pas seulement sur la problématique définitoire de la notion de compétence, jugée complexe par des auteurs tels que Coulet (2016), Kahn et

---

<sup>63</sup> Page non citée dans la source consultée.

<sup>64</sup> Page non citée dans la source consultée.

Rey (2016), Delory (2002)... Son évaluation suscite également un intérêt sans pareil étant donné que les organismes scolaires et sociaux réclament des outils assurant une évaluation authentique des compétences des évalués. Ces outils ne sauraient être conçus et mis en œuvre sans une précision de la notion de compétence en évaluation, qui ne pourrait être possible sans la prendre dans le contexte d'enseignement/apprentissage. Dans le cadre de l'APC, Roegiers (2000) distingue des compétences terminales, transversales et disciplinaires. Dérivées d'une macro-compétence (Objectif Terminal d'Intégration ou Éditions) qui renvoie au profil de sortie<sup>65</sup> attendu au terme d'un cycle, ces compétences n'occupent pas le même statut. Ainsi, les compétences transversales, peu évaluables et génériques, ne sont pas enseignées explicitement. Contrairement à ces compétences, les compétences disciplinaires constituent des objets d'enseignement explicite à travers les situations d'apprentissage pour doter les apprenants de moyens leur permettant d'atteindre des compétences terminales. Ce sont ces dernières que plusieurs chercheurs comme Gérard et Roegiers (1993), Le Boterf (1994), Perrenoud (1997) et Roegiers (2000) désignent par *compétences* tout court. Roegiers en dénombre les caractéristiques suivantes :

- Elle a un caractère complexe (mobilisation de plusieurs ressources S, SF, SE<sup>66</sup>) ;
- Elle doit pouvoir s'exercer de manière concrète ;
- On peut lui associer une famille de situations ;
- Elle caractérise un niveau donné de manière précise ;
- Elle est évaluable. (2006, p. 67).

De toutes ces compétences, une évaluation sommative porte essentiellement sur les compétences terminales que vise le programme annuel. Cette affirmation est cependant secouée par les travaux de certains chercheurs qui remettent en question le caractère évaluable de la compétence.

### 2.1.2. Peut-on évaluer les compétences ?

« *Il n'est pas évident d'évaluer les compétences des élèves* » (Gérard, 2009, p. 11) répondent certains. Cette citation qui reconnaît la difficulté d'une évaluation de compétences en admet toutefois la possibilité. Pourquoi est-il difficile d'évaluer les compétences des apprenants ?

---

<sup>65</sup>La notion de *profil de sortie* désigne « *les caractéristiques attendues de l'élève au terme d'un niveau d'études* » (Roegiers, 2006, p. 59). Elle s'oppose à celle de *profil d'entrée* qui renvoie à un aperçu descriptif des compétences de l'apprenant avant d'entreprendre un ensemble d'apprentissage.

<sup>66</sup> S : Savoir, SF : Savoir-Faire, SE : Savoir-Etre...

Selon Gérard (2009), les enseignants ne disposent pas d'outils évaluatifs pertinents pour réussir une telle évaluation : les renseignements sur les situations complexes, les critères et les indicateurs à introduire pour une évaluation réelle des compétences sont insuffisants pour une pratique enseignante efficace. Ces situations ne constituent pas cependant le seul outil évaluatif à intégrer dans l'univers scolaire dans la mesure où d'autres moyens tels que les portfolios ont donné les résultats escomptés dans certains milieux éducatifs, des outils que les enseignants ne sont pas assez préparés à investir. En effet, leur formation initiale et continue ne les forge pas suffisamment pour faire face à la complexité de leurs nouvelles tâches. Dans ce sens, Beckers affirme que l'actuelle approche, par rapport à la pédagogie par objectifs l'ayant précédée, correspond à une « *une modification dans l'ergonomie du métier d'enseignant à laquelle ceux-ci estiment souvent ne pas être préparés* » (2002, p. 148). Insistant indirectement sur cette faille de formation, Carlier et Parmentier (2002) attribuent les difficultés des enseignants ne proviennent pas seulement de la complexité de l'évaluation des compétences mais concernent la mise en œuvre de cette approche par les compétences.

Dans le cas où les enseignants ont certaines connaissances théoriques sur ces outils, ils ne peuvent les intégrer aisément pour deux principales raisons. La première est leur réticence au changement. En dépit de l'adoption de l'approche par les compétences qui a rénové les pratiques enseignantes et évaluatives, de nombreux enseignants perpétuent leurs pratiques pédagogiques, d'où la difficulté de recourir à de nouveaux outils évaluatifs. Cet attachement aux démarches relevant des anciennes pédagogies s'explique également par les modalités d'évaluation dans les examens officiels. Le système éducatif algérien illustre cette affirmation : comme les examens n'évaluent pas réellement les compétences ciblées par les programmes scolaires, les enseignants tendent à moduler leurs enseignements de façon à préparer leurs apprenants à réussir leurs examens (Gérard, 2006).

La seconde raison est relative aux ouvrages, documents et guides dont les enseignants ont besoin pour réguler leurs pratiques. Dans ce sens, peu de documents les initient aux outils, pratiques et activités à intégrer pour évaluer les compétences de leurs apprenants dans les disciplines concernées. Si ces documents existent, ils portent sur des publics particuliers, montrant des modalités pour évaluer des disciplines particulières, car destinés à des enseignants de systèmes éducatifs particuliers, d'où la difficulté de les appliquer directement dans des milieux distincts avec des publics différents. Pour illustrer nos propos, nous nous référons au système éducatif algérien, contexte de notre recherche. Bien que l'Approche Par les Compétences soit officiellement appliquée depuis 2003, nous assistons actuellement à une révision des programmes, des contenus, des manuels et des pratiques pour une meilleure

adaptation de cette approche à la particularité des apprenants algériens. Concernant les outils évaluatifs, nous comptons deux ouvrages collectifs qui explicitent les principes, caractéristiques, les objectifs...de cette approche tout en évoquant la modalité d'évaluation, d'une manière générale. Il s'agit de « *La refonte de la pédagogie en Algérie : défis et enjeux d'une société en mutation* » et « *Réforme de l'éducation et innovation pédagogique en Algérie* », dirigés par Toualbi-Thaâlibi et Tawil. Cette sécheresse éditoriale se justifie par le nombre restreint de pratiques évaluant les compétences, dû au caractère récent du thème.

Contrairement à ce premier point de vue de Gérard (2009) qui en reconnaissant la difficulté de l'évaluation des compétences en admet la possibilité, un second point de vue affirme que les compétences ne sont pas évaluables. De nombreux auteurs à l'instar de Riba et Mègre assurent que ce sont les performances que nous évaluons pour juger ultérieurement la compétence des évalués. Ils évoquent plutôt une évaluation des performances de compréhension et de production en langues vu que cette « *performance sera l'expression de la compétence dans un contexte déterminé* » (2014, p. 134). Cette évaluation de performances est elle-même contestée dans la mesure où le constat de la performance ne garantit pas la compétence vu que « *la première réalisation peut tenir à une convergence de circonstances favorables extérieures à l'individu, voire tenir au hasard* » (Kahn & Rey, 2016, p. 6). Or, la répétabilité de la performance ne prouve pas également la compétence, comme le montrent Kahn et Rey (2016), Soderstrom et Bjork (2015)... En dépit de ces réserves, de nombreux auteurs continuent à évoquer une évaluation de compétences.

Tel que nous l'avons explicité ci-dessus, l'argument des opposants à une évaluation de compétences découle de la dichotomie conceptuelle compétence/performance qu'il convient de dépasser pour réfléchir à des outils évaluatifs utiles. Cette utilité, nous la retrouvons dans la position de Roegiers (2005), un des précurseurs de l'APC dans le système éducatif algérien. Cet auteur assure la possibilité d'une évaluation des compétences réelles des apprenants, à des conditions que nous reprenons ci-dessous.

La première condition citée par Roegiers a trait à la formulation des compétences. Il avance que « *les compétences doivent être rédigées de manière à pouvoir être évaluées* » (2005, p. 109). De ce fait, l'énoncé de la compétence doit être précis exprimant la production concrète attendue des apprenants sans pour autant nuire à la complexité de la situation proposée. A cette première condition, il ajoute une deuxième relative au nombre des compétences qui doit être restreint : 2 à 3 compétences disciplinaires durant une année scolaire. Cette recommandation n'est pas fortuite étant donné qu'« *augmenter ce nombre présente un travers majeur : on risque de réduire la complexité, en découpant davantage les choses, et d'évaluer plus un savoir-faire*

*limité qu'une compétence* » (2005, p. 110). En effet, la multiplication de la compétence globale en plusieurs sous-compétences dé-complexifie la situation d'évaluation proposée. De cette manière, nous aboutissons à des exercices d'application évaluant plutôt le savoir-faire des apprenants qu'à des activités évaluant les compétences réelles des apprenants, ou leur capacité à mettre en œuvre leurs savoirs et savoir-faire dans une réalisation concrète. La dernière condition énumérée par Roegiers concerne les activités constitutives des épreuves d'évaluation. Ces activités consistent en des situations complexes, signifiantes et motivantes qui assurent la vérification de la compétence évaluée. Certes, nous ne pouvons pas prétendre à la possibilité d'en évaluer la maîtrise à travers des situations sociales authentiques, d'où le recours à des situations simulées qui reflètent le milieu social.

En considérant les trois conditions listées par Roegiers, nous les trouvons dépendantes de la compétence des enseignants. Ceci dit, une évaluation des compétences est possible pourvu que les enseignants disposent d'une maîtrise suffisante des outils assurant une telle tâche, ce qui rejoint la conception de Gérard (2009), évoquée supra. Théoriquement parlant, une évaluation des compétences est possible, mais, qu'en est-il de la pratique ? Évalue-t-on réellement des compétences dans le système éducatif algérien ? Les enseignants algériens sont-ils pédagogiquement suffisamment armés pour une adaptation aux changements introduits par l'approche par les compétences ?

A ces questions, les auteurs contributeurs dans les deux ouvrages collectifs -dirigés par Toualbi-Thaâlibi et Tawil- sur la réforme de la pédagogie en Algérie, à l'instar de Roegiers, Gérard... tentent de répondre de manière détaillée à travers leurs articles. Nous nous limitons à mentionner deux idées essentielles qui renforcent nos propos. La première est la réticence de certains enseignants au changement : habitués à des évaluations traditionnelles -constituées essentiellement d'exercices structuraux- ces enseignants se trouvent incapables d'introduire des pratiques évaluatives innovantes (Roegiers, 2005). La seconde est la déficience de leur formation en matière de la nouvelle approche, qui les a conduits à la maintenir même dans leurs pratiques enseignantes, comme l'explique Seghouani dans cette citation :

Inspecteurs et enseignants n'ayant pas bénéficié d'une formation de base sur la pratique d'une pédagogie de l'intégration à même de leur faciliter l'appropriation de l'approche préconisée par les nouveaux programmes d'enseignement, ont vite fait de se replonger dans "le confort" que leur procuraient leurs bonnes vieilles méthodes, aidés en cela par les contenus de nombreux nouveaux manuels qui étaient en décalage par rapport aux orientations des programmes. (2005, p. 74).

Nous devons préciser que des formations ont été organisées et dispensées aux inspecteurs et enseignants en vue de les orienter dans l'introduction des pratiques relevant de l'approche par les compétences. Insuffisantes ou infécondes, ces formations n'ont pas eu les effets escomptés étant donné que les pratiques traditionnelles subsistent en enseignement et en évaluation des apprentissages scolaires. Cette application boiteuse de l'APC n'est pas seulement due à la formation des enseignants, les manuels scolaires mis à leur disposition ne sont pas à la hauteur des attentes, notamment si de nombreux enseignants algériens demeurent esclaves des supports et des activités retenus. Les lacunes de ces manuels se confirment par les décisions de réécriture des manuels scolaires des divers cycles et paliers qui se poursuivent actuellement. Sans nous étaler sur les incorrections et les contradictions de toutes sortes soulevées par les inspecteurs, les enseignants, les apprenants et les divers acteurs pédagogiques et sociaux, nous affirmons que le système éducatif algérien a théoriquement adopté l'APC et qu'un long chemin vers une application réelle reste à parcourir par tous les acteurs.

La réalité de l'évaluation des compétences dans le système éducatif algérien nous pousse à nous lancer dans cette étude en vue d'aboutir à des dispositifs d'évaluation sommative qui évaluent réellement et fidèlement les compétences de nos apprenants. Nous admettons la possibilité d'une telle évaluation mais nous reconnaissons également le manque d'outils assurant la réussite de leur tâche. Néanmoins, toute entreprise de conception d'outils d'évaluation doit être précédée de déterminations sur les compétences évaluables. Quelles compétences faudrait-il alors évaluer en classe de FLE ? Avant de répondre à cette question, qui sera abordée dans la deuxième phase de l'élaboration de dispositif d'évaluation, nous devons préciser si une évaluation de compétences exclut celle des ressources.

### **2.1.3. Peut-on évaluer les ressources dans toute évaluation scolaire ?**

Cette question nous semble importante à éclaircir dans le contexte de notre étude. En effet, l'évaluation scolaire -dans une APC- porte sur deux objets essentiels : ressources et compétences. Si les compétences sont à évaluer dans cette approche, une évaluation de ressources n'est pas exclue. Selon Gérard (2009), une évaluation scolaire doit alterner une évaluation de compétences et de ressources, les deux éléments étant objets d'enseignement/apprentissage.

Une évaluation de ressources porte sur les savoirs et les savoir-faire que les apprenants mobilisent pour résoudre les situations-problèmes ultérieurement proposées. Elle s'opère à travers des exercices d'application ou des activités de QCM, appariement... Souvent, les auteurs la préconisent durant les apprentissages à des fins de régulation, comme l'affirme



Gérard (2009). De cette manière, une telle évaluation relève d'une évaluation formative préparant la réussite des apprenants dans une évaluation sommative, évaluant essentiellement leurs compétences. Celle-ci, contrairement à la première, évalue les compétences visées par l'enseignement/apprentissage à travers des situations complexes où les ressources sont mobilisées. Autrement dit, elle ne s'opère pas à travers des items isolés évaluant la maîtrise des savoirs et des savoir-faire inculqués aux apprenants. Elle est effectuée au terme de l'action pédagogique qui ne se réduit pas à un enseignement des ressources<sup>67</sup>, mais porte également sur un enseignement de leur intégration.

Cependant, selon Gérard, une évaluation sommative, un concept que nous préférons - pour son aspect global- à celui d'évaluation certificative que l'auteur emploie, n'exclut pas entièrement une évaluation de ressources. Dans ce sens, cet auteur précise que « *certaines ressources importantes peuvent aussi faire l'objet d'évaluations certificatives, mais l'évaluation des ressources ne devrait pas dépasser 25% des points de l'évaluation certificative* » (2009, p. 51). Cette citation admet certainement la possibilité d'une évaluation de ressources dans une évaluation sommative, mais elle ne passe pas inaperçue étant donné qu'elle nous inspire des questions extrêmement importantes : quelles sont ces ressources importantes que les enseignants évaluent ? Sont-elles évaluables dans toutes les disciplines ?

Toutes les questions qui étoffent notre travail démontrent en réalité la complexité de la problématique de l'évaluation des compétences. En effet, chaque réponse entraîne des questions dont les réponses n'assouissent aucunement les interrogations des enseignants et évaluateurs. En précisant que les compétences sont évaluables, et que les ressources le sont également, la question des outils assurant une telle évaluation s'impose. Les enseignants ont besoin de moyens pratiques leur permettant d'évaluer ces compétences développées chez leurs apprenants.

## **2.2. Des précisions sur les examens sommatifs à concevoir**

Lors de la planification des examens, nous devons distinguer deux cas qui requièrent des mesures particulières. Dans le premier cas (les examens internes), les épreuves sont conçues par les enseignants chargés de l'enseignement des matières concernées par l'évaluation. Dans ce cas-ci, certaines précisions (les moments de l'évaluation, les caractéristiques des évalués...) que nous aborderons dans ce qui suit ne revêtent pas une importance majeure. Dans le second

---

<sup>67</sup> Pour ce qui est des ressources, la distinction établie par Gérard entre les ressources internes et les ressources externes est déjà évoquée.

cas (les examens externes), que nous ne négligeons pas entièrement, les examinateurs qui sont chargés de la planification et de la conception des examens sont distincts des enseignants. Ce qui rend nécessaire toutes les précisions assurant une évaluation rigoureuse des compétences des apprenants. Quelles précisions la planification devrait-elle prendre en considération ?

### **2.2.1. Le moment, nombre et durée des examens**

Dans le cas des examens externes, la date de passation est à communiquer aux concepteurs pour la finalisation des examens à concevoir dans les délais. D'ailleurs, les concepteurs doivent être au courant de l'état d'avancement dans les programmes et les contenus proposés afin que des activités soient conçues tenant compte des contenus réellement enseignés.

S'agissant des examens scolaires, souvent élaborés par les enseignants chargés de la matière, un canevas fixant les dates des examens programmés durant une année scolaire est communiqué. En fonction de la date prévue pour l'examen, la conception des activités évaluatives, leur révision... sont planifiées, dans des moments précis. Mentionnons que la conception des examens tient également compte de la durée que les enseignants leur consacrent. Cette information est précieuse dans le contexte évaluatif étant donné que le nombre des questions à prévoir et la complexité des situations en dépendent. Si la durée de l'examen est courte, les enseignants ne pourront proposer un nombre élevé d'items à résoudre ou de situations complexes qui mobilisent des opérations cognitives de haut niveau.

Le canevas fixant les dates des évaluations sommatives en précise également le nombre. Cette information, sans importance de prime abord, est extrêmement utile car elle détermine des choix tels que la répartition du programme. Ainsi, selon le nombre des examens programmés, les contenus sont répartis de façon à ce que tous les éléments enseignés soient évalués. Il est important de souligner à ce niveau que les enseignants ont la possibilité d'apporter des modifications jugées nécessaires au nombre et aux dates des examens prévus.

Le nombre d'examens est, aux yeux des enseignants et de leurs apprenants, un indicateur de leur importance. Plus leur nombre augmente, plus leur valeur diminue, d'où la nécessité de bien étudier le nombre d'évaluations sommatives à introduire au cours d'une année scolaire. Un nombre élevé d'examens a des retombées indésirables sur les enseignants et sur leurs apprenants. Pour les premiers, ils seront plus occupés à préparer des examens qu'à planifier leurs cours, suivre le développement des apprenants et remédier à leurs lacunes. De cette manière, les enseignants se voient réduits à des concepteurs d'examens et l'évaluation n'assure pas les fonctions régulatrices qui lui sont associées. Quant aux apprenants, une évaluation répétitive constitue une source d'instabilité, de stress et de peur, des facteurs qui entravent une

bonne assimilation des contenus programmés. Tenant compte de ces données, le calendrier des examens à proposer aux enseignants devrait être conçu par des experts de façon à ne pas fatiguer les apprenants ou les décourager.

### **2.2.2. Norme à fixer**

Cet élément est d'une importance cruciale dans toute évaluation. Par norme, nous entendons le seuil de réussite fixé dans un examen donné. Dans les examens externes, la norme requise serait de donner des réponses correctes à toutes les questions posées (matières scientifiques). Qu'en est-il des évaluations scolaires ?

Dans des démarches classiques, un classement qui compare le rendement des apprenants est établi. Les derniers de ce classement -dont le nombre est déterminé par les responsables (le directeur et les enseignants)- échouent car la norme requise est de ne pas être dans la catégorie des apprenants de la queue du classement. Cette norme pourrait être contestée dans la mesure où elle nuirait à l'équité en évaluation. Certains apprenants qui échouent dans un groupe pourraient avoir des compétences meilleures que celles des apprenants qui réussissent dans leurs groupes. Déterminer un niveau de réussite valable pour tous constitue ainsi un moyen pour une évaluation correcte et des décisions justes.

C'est la raison pour laquelle il est nécessaire de fixer un seuil de réussite déterminant ce qui est requis pour déclarer tel apprenant compétent. Dans de nombreux systèmes éducatifs, comme le système éducatif algérien, où la notation est adoptée, la norme de réussite est la note 10 : les apprenants qui l'obtiennent sont déclarés admis pour passer à un niveau supérieur. Cette norme de réussite est déterminée par les organismes ayant pour mission la conception des programmes, des contenus et des modalités d'évaluation.

### **2.2.3. Communication des résultats : quels outils ?**

Les résultats que les apprenants obtiennent lors des évaluations sommatives trimestriels sont communiqués aux acteurs impliqués dans ce processus évaluatif. En effet, ils doivent être communiqués aux parents des élèves et au personnel de l'institution par différents moyens. Ce sont ces moyens que les enseignants doivent préciser lors de la phase de planification.

De ces moyens, les enseignants recourent souvent aux notes, bulletins..., des outils fréquents, mais souvent contestés par les praticiens en raison de leurs limites informatives. Nous ne nous étalons pas sur ces outils qui seront explicités d'une manière détaillée dans les éléments suivants de notre étude. Outre ces outils, la phase de planification pourrait programmer des rencontres entre enseignants et parents pour étudier les progrès des apprenants, analyser leurs résultats et proposer des solutions aux dysfonctionnements constatés.

#### **2.2.4. Conditions de passation à établir**

Ces conditions sont liées à des mesures que les enseignants discutent lors de la planification pour assurer le bon déroulement des épreuves. Les enseignants doivent en premier lieu s'adapter au matériel et aux moyens dont disposent leurs établissements. Dans des établissements intégrant largement les outils technologiques, les examens pourraient être proposés en version électronique, les examens oraux pourraient être intégrés... Il en est de même pour les moyens financiers qui déterminent la qualité du papier fourni, les couleurs dans les images ou illustrations insérées... En évoquant la qualité du papier, nous entendons certes les examens proposés sous format papier. Ces examens requièrent plusieurs critères pour en assurer le bon déroulement. Entre autres, nous citons : la taille de l'écriture, la lisibilité, la clarté des caractères... En plus du matériel et des moyens financiers, les conditions de passation prennent en considération le lieu de passation qui doit être organisé de façon à répartir les apprenants sur les salles en tenant compte de leur aménagement. Ces éléments, qui pourraient sembler superflus aux lecteurs, ne le sont pas si nous prenons en considération les obstacles qui jaillissent lors de la passation des examens tels que la non-lisibilité des textes proposés, le manque de places... Ces obstacles, de conséquences variées, biaisent considérablement le rendement des apprenants aboutissant ainsi à une évaluation boiteuse de leurs compétences.

Les éléments, listés supra, que la phase de planification doit absolument traiter ne sont pas exhaustifs : nous en citons ceux liés à une évaluation interne. En effet, d'autres éléments sont considérés cruciaux lors de la planification comme la coordination entre les enseignants assurant la même matière ainsi que la coordination avec d'autres établissements, la répartition des tâches, la conception de sujets-réserves pour éviter certains imprévus... De ces éléments, nous n'abordons pas de manière détaillée que les caractéristiques des apprenants et leur nombre : un élément clé selon le « *Manuel pour l'élaboration et la passation de tests et d'examens de langue* ». Il est connu que le nombre des apprenants détermine le nombre d'enseignants surveillant l'examen et corrigeant les copies. Il en est de même pour leurs caractéristiques (âge, niveau...) qui conditionnent les thèmes sélectionnés, les supports choisis... En dépit de son importance dans une évaluation externe, cet élément ne revêt pas la même valeur dans une évaluation interne où les enseignants évaluent des apprenants dont ils n'ignorent pas les caractéristiques et le nombre. Avec l'achèvement de cette phase de planification, dont la réussite assure celle des étapes suivantes, une nouvelle phase de l'élaboration de dispositif évaluatif débute : la conception.

### **3. Conception des instruments de mesure et leur expérimentation**

Suite à la planification du dispositif, les enseignants entament la conception des outils évaluatifs. Certains auteurs la désignent également par *prise d'informations* dans la mesure où « *l'enseignant construit ses instruments d'évaluation et les soumet à ses étudiants (Aubin, 2015 ; Leroux et Bélair, 2015) afin d'obtenir des informations pour la prochaine étape, le jugement (Aubin, 2015)* » (Roy, 2017, p. 24). Autrement dit, cette phase -à travers les activités évaluatives conçues puis soumises aux apprenants- assure la collecte des données permettant aux enseignants d'émettre leur jugement. Mentionnons que nous préférons le concept *conception* à celui de *prise d'information*, que nous jugeons restrictif car il insiste sur un moment particulier de cette phase (la collecte des informations) en omettant des moments aussi importants (l'élaboration des activités). Quels moments cette phase englobe-t-elle ?

### **3.1. Construction des outils de mesure pour une classe de FLE**

Le premier moment de la conception est la construction des outils de mesure qui permettent d'évaluer les compétences des apprenants. Pour ce faire, il est nécessaire de déterminer la formulation de ces compétences en objectifs dans les programmes scolaires proposés. Le CECRL (2011) distingue trois objectifs essentiels -étroitement liés au niveau des apprenants- visés par l'enseignement des langues. Ainsi, les objectifs liés à la compétence générale sont visés avec des apprenants débutants, des objectifs en rapport avec la compétence communicative langagière sont développés chez les apprenants de niveau intermédiaire, alors que des objectifs en rapport avec les activités langagières spécifiques sont souvent tracés avec des apprenants adultes. Quelle catégorie d'objectifs devons-nous retenir ? Et comment pouvons-nous les évaluer ?

#### **3.1.1. Que faut-il évaluer ?**

La première catégorie d'objectifs, les compétences générales individuelles, englobe les savoirs<sup>68</sup>, savoir-faire<sup>69</sup>, savoir-être<sup>70</sup> et savoir-apprendre<sup>71</sup>. De ces quatre types de savoirs, la littérature scientifique consultée n'aborde que l'évaluation des deux premiers : savoirs et savoir-faire. Hadji (2015) explique que l'évaluation des savoirs dans le milieu scolaire s'effectue à travers des activités de restitution alors que celle des savoir-faire nécessite des activités d'application. Pour la seconde catégorie (les compétences communicatives langagières), elle

---

<sup>68</sup>Les savoirs sont des connaissances déclaratives, très valorisées dans l'univers scolaire (Vianin, 2009).

<sup>69</sup>Les savoir-faire désignent la capacité d'appliquer les savoirs déclaratifs dans des contextes.

<sup>70</sup> Le savoir-être relève des caractéristiques individuelles ou attitudes manifestées dans les interactions sociales.

<sup>71</sup>Le savoir-apprendre est la mobilisation des savoirs, savoir-faire et savoir-être pour la découverte de ce qui est inconnu.

est constituée de trois composantes essentielles : une composante linguistique<sup>72</sup>, une composante sociolinguistique<sup>73</sup> et une composante pragmatique<sup>74</sup>. Ces compétences communicatives sont évaluées à travers des activités de restitution, d'application, de réinvestissement, des situations... La dernière catégorie, c'est-à-dire les activités langagières, est répartie en activités de réception, production, médiation<sup>75</sup> et interaction<sup>76</sup>, tant à l'oral qu'à l'écrit. De toutes ces activités, celles de réception et de production sont les plus basiques et fondamentales car l'interaction et la médiation dépendent de leur maîtrise. Ces activités langagières, qui s'inscrivent dans des domaines variés de la vie personnelle, scolaire ou professionnelle, s'entrecroisent étant donné que la réception et la production sont présentes dans l'interaction et la médiation.

De ces trois catégories, la dernière a particulièrement inspiré les concepteurs des examens dans les univers scolaires et des tests de grande envergure (DELFI, DALF...). Nous y évaluons séparément deux activités de réception et de production, tant à l'oral qu'à l'écrit, à travers des activités diverses. Cette évaluation séparée des compétences (compréhension et production), qui relève du schéma classique de l'évaluation selon Huver et Springer (2011), n'est plus actuellement admise dans la mesure où l'APC privilégie les situations en tant qu'instruments de mesure qui évaluent indissociablement les deux compétences citées. Cette nouvelle tendance s'explique par une réalité incontestable en évaluation de langue : l'impossibilité de séparer les deux compétences. En effet, certaines activités qui évaluent la compréhension (les questions ouvertes) exigent la formulation des réponses, ce qui mobilise la compétence expressive car les apprenants doivent être en mesure de rédiger des réponses reflétant avec exactitude leurs pensées. Le même principe s'applique sur l'évaluation de la production étant donné que la compréhension de la situation s'impose pour que les apprenants rédigent des énoncés cohérents. Vu que l'évaluation indépendante des compétences relève d'un modèle classique, non conforme aux exigences des approches récentes, il devient

---

<sup>72</sup> La composante linguistique se rapporte aux connaissances relatives aux dimensions du système linguistique (lexique, syntaxe, conjugaison...), indépendamment des situations sociales de leur usage.

<sup>73</sup> La composante sociolinguistique vise à déterminer les paramètres socioculturels qui régissent l'emploi des structures linguistiques assimilées par les apprenants.

<sup>74</sup> La composante pragmatique regroupe l'emploi des connaissances linguistiques dans des énoncés cohérents adaptés à la situation communicative.

<sup>75</sup> La médiation consiste en une production orale ou écrite par laquelle le locuteur joue le rôle d'intermédiaire pour assurer l'intercompréhension entre deux ou plusieurs interlocuteurs. Parmi les activités de médiation, le CECRL (2011) recense l'interprétation, la traduction, le résumé et la reformulation.

<sup>76</sup> Les interactions consistent en des échanges entre deux ou plusieurs interlocuteurs. De cette manière, elles alternent les deux activités précédemment citées : réception et production.

nécessaire -dans un souci de validité- de recourir à des instruments évaluant indissociablement la compréhension et la production. Quels sont les instruments qui assurent une telle évaluation ?

### **3.1.2. Évaluer par des situations**

Pour une évaluation valide des compétences des apprenants, des auteurs tels que Roy (2017), Bélair (2014) et Leroux et Bélair (2015) préconisent le recours à des tâches complexes qui sont ancrées dans des situations. Ces situations constituent ainsi des instruments permettant une évaluation des compétences de leurs apprenants, à travers les performances manifestées en réponse aux tâches proposées.

## **3.2. La situation d'évaluation**

L'une des notions clés introduites par l'APC est celle de situation. En effet, les apprentissages sont présentés puis évalués dans des situations : les premières sont désignées par situations d'apprentissage alors que les secondes sont appelées situations d'évaluation. En quoi consistent-elles ?

Semblables aux situations d'intégration (Gérard, 2009), les situations d'évaluation assurent l'évaluation du degré de maîtrise des compétences développées chez les apprenants à travers les situations d'apprentissage. Pour les définir, nous reprenons la définition proposée par Parent. La situation d'intégration y est conçue comme « *une situation complexe, contextualisée et signifiante qui engage les apprenants à utiliser tous ses savoirs, issus de plusieurs cours et qui nécessite la réalisation d'une production clairement précisée* » (Parent, 2008, p. 118). Au-delà des caractéristiques de cette situation listées par Parent, la citation met l'accent sur la fonction des situations d'intégration qui assurent aux apprenants le transfert de ce qui est appris dans des productions concrètes, répondant à des contextes particuliers. Ce sont les contextes sélectionnés qui permettent aux acquis de se concrétiser, donnant aux enseignants un aperçu sur le degré de leur maîtrise.

En dépit de la similitude entre les deux, les situations d'intégration et d'évaluation ne sont pas à confondre en raison de leurs objectifs distincts : les premières visent un apprentissage –à travers l'investissement de ce qui est appris dans un contexte donné-, alors que les secondes assurent une évaluation de la maîtrise des compétences développées chez les apprenants.

### **3.2.1. Caractéristiques des situations d'évaluation**

Les situations évaluatives ont des caractéristiques qui varient d'un auteur à un autre. Ainsi, Parent (2008) en liste trois : les situations sont complexes, contextualisées et signifiantes. Quant à Lemenu et Heinen (2015), elles évoquent la complexité et la représentativité des situations sélectionnées des situations réelles, autrement dit l'authenticité. Ces caractéristiques,

que de nombreux auteurs tels que Rey (2017), Roegiers (2005, 2006), Gérard (2009)... abordent dans leurs écrits sont traitées dans les éléments suivants.

- La complexité : une caractéristique principale

De toutes les caractéristiques des situations, la complexité est la plus abordée par les auteurs qui s'accordent à affirmer la nécessité de recourir à des situations complexes pour une évaluation des compétences (Roegiers, 2006 ; Carette, 2009). Qu'est-ce qu'une situation complexe alors ?

Pour définir les situations complexes, Gérard (2008) se réfère à Roegiers (2003) qui oppose une *situation compliquée* à une *situation complexe*. Contrairement à la première qui intègre des savoirs difficiles (nouveaux, inconnus ou non-maîtrisés) pour les apprenants, la seconde intègre des savoirs maîtrisés, investis et articulés différemment dans les situations proposées. Selon Roegiers (2003), la complexité ne relève pas de la nature des éléments investis pour résoudre un problème donné étant donné que ces éléments ou ressources sont déjà développés chez les apprenants. Elle relève plutôt de la combinaison de ces éléments dans les situations nouvelles : les apprenants les mobilisent en tenant compte de la particularité des contextes proposés. Rejoignant cette conception, Roegiers (2006) précise que la complexité se situe au niveau de l'articulation des ressources apprises. Selon lui, complexe « *signifie qu'il y a articulation en contexte d'un ensemble d'éléments, pas nécessairement difficiles en eux-mêmes, mais c'est cette articulation même qui constitue un défi pour l'élève* » (2006, p. 67).

Tenons à préciser la difficulté d'une évaluation à travers des situations complexes, notamment en raison des exigences de leur conception et correction. En dépit de cette difficulté, de nombreux auteurs tels que Gérard (2008, 2009), conscients de l'apport des situations complexes en évaluation scolaire, préconisent leur investissement pour une évaluation des compétences des apprenants. D'ailleurs, pour cet auteur, la situation complexe est le seul instrument qui assure une évaluation des compétences. Il affirme que la situation complexe

Nécessite une véritable compétence pour pouvoir la résoudre, celle d'analyser la situation, d'identifier et de mobiliser des ressources pertinentes- qui peuvent être différentes selon les élèves- et de les articuler entre elles de manière inédite, ponctuelle et éphémère et souvent propre à chaque individu (Gérard, 2005, p. 2).

Les définitions de la complexité suggèrent une deuxième caractéristique des situations évaluatives : le caractère inédit. En présentant la complexité comme une articulation de ressources connues pour répondre aux exigences de nouveaux contextes, les auteurs (Gérard, 2008 ; Roegiers, 2003...) l'associent à l'aspect inédit. Une situation nouvelle ou inédite



mobilise des ressources maîtrisées en les articulant différemment. C'est la raison pour laquelle de nombreux auteurs juxtaposent les qualificatifs complexe et inédit en abordant les caractéristiques des situations d'évaluation. Ainsi, Rey (2012) évoque la nécessité de proposer des situations complexes et inédites. Il aborde le concept de *complexité inédite* que des auteurs tels que Crahay (2006) objectent en évaluation scolaire, la trouvant génératrice d'injustices. Notons que le niveau de complexité des situations dépend de celui des apprenants : une situation complexe pour des apprenants débutants ne l'est pas pour des apprenants de niveau intermédiaire. D'ailleurs, les ressources que les apprenants investissent se complexifient : les compétences deviennent ressources à un niveau plus avancé (Roegiers, 2006).

Grâce aux situations complexes, une évaluation ne se réduit plus à évaluer la maîtrise de ces ressources. Elle donne un aperçu sur la capacité qu'ont les apprenants à articuler leurs apprentissages dans une production adéquate au contexte proposé, souvent similaire à un contexte authentique. En évoquant le contexte, nous arrivons à une deuxième caractéristique de la situation d'évaluation qui doit être contextualisée. En quoi consiste une situation contextualisée ?

- Situations contextualisées

De prime abord, parler d'une situation contextualisée ne semble pas justifiée dans la mesure où les notions *situation* et *contexte* sont communément confondues. L'usage commun les emploie indifféremment pour évoquer les circonstances dans lesquelles un évènement donné s'inscrit. Cependant, les deux notions renvoient à des réalités distinctes que de nombreux auteurs (Piron, 2013 ; Franckel, 2006 ; Cuq, 2003 ; Reuter, 2011 ; Coste et Galisson, 1976...) explicitent dans leurs définitions.

De toutes ces définitions, nous retenons celle de Reuter, plus pertinente pour notre étude, qui introduit la notion *contexte de la situation*. Cette introduction est révélatrice du point de vue de l'auteur qui ne conçoit plus le contexte et la situation comme des termes opposés. Le contexte est, pour lui, une partie intégrante de la situation. Reuter formule la définition suivante :

La notion de contexte de la situation rassemble (...) les éléments de la situation avec lesquels le sujet est dans une relation de sens. Plus précisément, il est l'ensemble des significations que le sujet attribue à ces éléments. Comme l'écrit Jean-Paul Bernié (2005 : 145), le contexte « englobe tout ce qui fait sens dans l'univers des tâches (2011, p. 53).

Le contexte est ainsi lié aux éléments linguistiques de la situation permettant d'interpréter ou de produire des énoncés de divers types, ainsi qu'aux éléments extralinguistiques (illustrations, tableaux, cartes, figures...). Ces éléments linguistiques et

extralinguistiques regroupés déterminent et conditionnent les productions des apprenants (Reuter, 2011). Notons que le contexte revêt une importance majeure dans la résolution de la situation-problème proposée. Dans ce contexte, Rey estime qu'il « *n'est pas qu'un prétexte à la vérification des connaissances, mais fait partie intégrante de la solution* » (2012, p. 2), d'où la nécessité de proposer aux apprenants des situations aux contextes clairs et explicites assurant une évaluation réelle de leurs compétences.

- Situations significatives

Pour préciser en quoi consiste une situation significative, nous nous référons à Gérard et Roegiers qui avancent que « *par situation significative, il faut entendre une situation dans laquelle l'élève peut être amené à utiliser seul l'objet d'apprentissage* » (2009, p. 212). Cette définition que les auteurs proposent insiste sur l'investissement autonome par les apprenants de ce qui est appris. Cette autonomie n'est cependant pas due à la situation elle-même dans la mesure où elle pourrait résulter d'une maîtrise satisfaisante des objets appris. Que les apprenants soient capables de transférer efficacement leurs apprentissages, sans guidage, ne relève pas des particularités de la situation. Cependant, Roegiers (2006) précise les éléments qui rendent la situation (qui doit être vraisemblable, illustrée et authentique) significative. Celle-ci s'inscrit dans un contexte qui motive les apprenants les amenant à un investissement de ce qui est appris.

Ces définitions, nous les renforçons par celle formulée par Dany. Selon lui, nous considérons significative toute situation qui « *du sens ou qui permet à l'élève de construire du sens à son propos* » (2013, p. 136). Cet aspect significatif provient du choix des thèmes qui relèvent des intérêts des apprenants, des situations ancrés dans leur vécu et dans leur quotidien... En effet, ce sont les situations que les apprenants confrontent souvent qui donnent du sens aux apprentissages programmés, dont la valeur dépend du potentiel d'investissement dans des situations réelles.

Cependant, cette caractéristique présente des limites indéniables aux enseignants-concepteurs de situations. Ceux-ci sont contraints de concevoir des situations significatives tout en tenant compte de l'hétérogénéité de leurs apprenants. Cette tâche est difficile dans la mesure où l'enseignant ne possède pas des pouvoirs magiques lui permettant de deviner ce qui pourrait avoir du sens aux yeux de ses apprenants. En plus de cette première difficulté, les enseignants sont confrontés à une deuxième : celle de devoir satisfaire la diversité des goûts, des intérêts... de leurs apprenants. Comme une situation significative pour certains ne l'est pas pour d'autres, concevoir une situation réellement significative devient une tâche illusoire.

- L'authenticité des situations

L'évaluation des compétences requiert le recours à des situations authentiques, selon Lemenu et Heinen (2015). Ces auteurs citent Prégent, Bernard et Kozanitis (2009) qui affirment que, pour évaluer les compétences, il est préférable de proposer des situations authentiques (sur le terrain). Comme une telle évaluation ne serait pas aisément menée dans un contexte scolaire, les enseignants recourent dans leurs pratiques évaluatives à des situations authentiques, autrement dit « *semblables à celles que l'on retrouve dans la réalité de la vie professionnelle ou de la vie quotidienne* » (Prégent, Bernard & Kozanitis, 2009, p. 139). Notons que la notion de situation authentique n'est pas récente. Elle remonte aux travaux de Wiggins (1998) et à ceux de Frenay et Bédard (2004). En effet, les travaux de Wiggins (1998) listent les caractéristiques d'une situation d'évaluation authentique. Nous les reprenons dans ces deux points essentiels :

- Des tâches réalistes et inédites : dans les situations authentiques, l'enseignant doit sélectionner des tâches similaires à celles de la vie authentique de ses apprenants, lui permettant de prédire leurs performances dans les situations sociales. En plus de ce premier critère, les tâches retenues devront être inédites. Le qualificatif *inédit* doit être précisé car il ne suppose pas la nouveauté absolue de la tâche, qui tout en étant similaire aux situations précédentes requiert une articulation innovée des ressources antérieurement apprises. Ces deux critères des tâches scolaires tout en assurant une évaluation réelle des compétences des apprenants assurent la régulation de tous les dysfonctionnements constatés, aboutissant de cette manière à une amélioration du produit réel.
- Mobilisation des opérations cognitives supérieures : contrairement aux activités évaluatives traditionnelles qui mobilisent les opérations cognitives de niveau inférieur (récitation, application...), les situations complexes et authentiques requièrent des apprenants une analyse et une réflexion pour résoudre la situation-problème retenue.

Admettons que des situations authentiques en évaluation scolaire ne pourraient être intégrées dans la mesure où elles exigent une évaluation des compétences des apprenants dans une discipline donnée en contexte réel. De cette manière, évaluer la compétence communicative des apprenants en FLE ne serait possible qu'en plaçant les apprenants dans des situations où ils seront contraints de communiquer en langue étrangère. En raison de contraintes institutionnelles et sociales, cette authenticité des situations est garantie en classe de FLE à travers des situations simulées, à travers desquelles les enseignants tentent de reproduire les propriétés des situations sociales.

En somme, de nombreuses caractéristiques sont requises pour que les situations évaluatives conçues soient conformes aux exigences pédagogiques et institutionnelles. Nous ne prétendons pas recenser toutes leurs caractéristiques que des auteurs tels que Roegiers (2006) listent dans leurs travaux.

- Des situations diverses aux paramètres identiques...

Qu'elles visent un apprentissage, une intégration ou une évaluation, les situations à proposer par les enseignants devront appartenir à une même famille de situations-problèmes. En quoi consiste une famille de situations ? Et quels sont les critères qui la définissent ?

Pour définir la notion de **famille de situations**, nous nous référons à Miled qui la conçoit comme « *ensemble de situations ayant un niveau de difficulté équivalent mais qui se rattachent à une même compétence* » (2005, p. 131). Cette citation insiste sur le niveau de difficulté identique dans les situations de même famille, sans pour autant déterminer les critères qui fondent leur cette famille. En effet, toute situation particulière comporte des éléments qui l'attachent aux autres situations intégrées à des fins d'apprentissage ou d'évaluation : ce sont les *invariants* ou les *paramètres*. En quoi consistent-ils ?

De nombreux auteurs, à l'instar de De Ketele (2013), s'accordent à définir les **invariants** comme des éléments communs aux situations appartenant à une famille précise, comme il est mentionné ci-dessus. Rejoignant la même conception, Lemenu et Heinen formulent la définition suivante des paramètres : « *les invariants sont les ingrédients de base, les caractéristiques communes incontournables de toutes les situations qui permettent de dire que celles-ci seront d'un niveau plus ou moins équivalent* » (2015, p. 100). Ces constituants basiques établissent une équivalence entre les situations assurant par là leur appartenance à une même famille. Définir ces invariants ou paramètres est d'une importance majeure vu leur rôle dans la conception de situations équivalentes aux situations proposées. Conscients de cette importance, nombreux sont les auteurs (De Ketele, 2013 ; Lemenu & Heinen, 2015) qui les énumèrent dans leurs écrits. Il s'agit des :

- Ressources à mobiliser : ces savoirs, savoir-faire et savoir-être que les apprenants mobilisent dans la situation évaluative sont ceux développés chez les apprenants à travers les situations d'apprentissage et d'intégration.
- Type de situation : le domaine dont il est question (vie quotidienne, environnement...) dans les diverses situations doit être identique : ce que les enseignants proposent en évaluation ne diffère pas de ce qui est proposé en cours.
- Type et nombre de supports : par supports, les auteurs désignent les documents que la situation regroupe (textes, photos...). Dans les situations correspondant à une

même famille, le nombre des supports ainsi que leur type doit être étudié : si la situation (A) comporte un texte court, la situation (B) doit comporter un seul texte de longueur identique.

- Type de tâche attendue : les tâches attendues, ou ce que les apprenants doivent réaliser, doivent être identiques dans les diverses situations de même famille (type textuel à produire...).
- Conditions : il est important que les conditions de réalisation des tâches retenues telles que la durée, la forme du travail, le matériel...se ressemblent.
- Critères d'évaluation : lors de la conception des situations évaluatives, les enseignants déterminent les critères minimaux et de perfectionnement qui les guident lors de la correction des copies de leurs apprenants. Souvent, ces critères sont stables dans une famille de situations...

Ces paramètres que les auteurs précisent assurent une équivalence des situations retenues dans une séquence pédagogique. Le recours au terme *invariants* pour les désigner est justifié dans la mesure où il insiste sur les éléments qui ne varient pas, c'est-à-dire stables, dans toutes les situations de la même famille. Contrairement à ces invariants, certains éléments sont particuliers à une situation pour en assurer l'aspect inédit. Cette distinction entre les invariants et les éléments qui varient nous conduit à évoquer deux notions-clés : le **paramétrage** et l'**habillage**. En quoi consiste chacune de ces notions ?

En vue de définir le paramétrage, De Ketele et Gérard (2005) l'opposent à l'habillage. Par opposition au premier qui se rapporte à une famille de situations, le second est relatif à une situation particulière : le paramétrage se rapporte aux principaux éléments de la situation, tandis que l'habillage concerne les éléments périphériques et particuliers de la situation tels que le contexte spécifique, les supports intégrés, les données concrètes... (De Ketele & Gérard, 2005). L'habillage assure ainsi le caractère inédit des situations : chaque situation en dépit des ressemblances avec les autres est inédite.

De ces deux notions, celle du paramétrage revêt une importance indéniable dans la mesure où elle assure une certaine standardisation des situations-problèmes, les reliant toutes à la compétence terminale, ce qui assure la validité et l'équité de l'évaluation scolaire des diverses disciplines. En se référant aux paramètres déterminés, les enseignants conçoivent des situations d'enseignement pour développer des compétences précises chez leurs apprenants, évaluées ultérieurement à travers des situations évaluatives respectant les mêmes paramètres. D'ailleurs, étant donné que la compétence ne pourrait être évaluée à travers une seule situation, le recours à plusieurs situations devient nécessaire pour émettre un jugement fondé et valide sur la

performance évaluée. Cependant, ces situations-ci doivent présenter des caractéristiques communes, d'où le recours aux paramètres. La maîtrise de ces paramètres devient une compétence nécessaire chez les enseignants, d'où la nécessité de les former à la conception de situations-problèmes conformes aux exigences. Dans ce contexte, des auteurs tels que Gérard (2009) proposent des grilles permettant aux enseignants de vérifier si les situations évaluative retenues sont idéales<sup>77</sup>. Tel qu'employé par Gérard (2009), le concept idéal ne renvoie pas au parfait inexistant : il désigne une situation qui présente toutes les caractéristiques nécessaires. Une situation idéale, qui présente de nombreuses caractéristiques évoquées supra, doit être également complète. Quand une situation évaluative est-elle jugée complète ?

### 3.2.2. Composantes des situations d'évaluation complètes

Toute situation, qu'elle vise un apprentissage, une intégration ou une évaluation, est constituée de trois composantes essentielles : support, tâche et consigne. En énumérant les éléments qui constituent toute situation-problème, Gérard (2009) ajoute aux trois composantes citées le contexte et la fonction. En effet, l'auteur repère le *contexte* (décrivant l'environnement), les *supports* (fournissant des informations contribuant à la résolution de la situation), la *fonction* (ou but), la *tâche* (ce qui est attendu) et la *consigne* (instruction). Comme le *contexte* (qui englobe la *fonction*) est précédemment expliqué, nous nous limitons à expliquer dans les éléments qui suivent les trois composantes essentielles :

#### 3.2.2.1. Les supports

Le premier constituant indispensable de toute situation d'intégration ou d'évaluation est le support. Ce constituant revêt une importance majeure dans la mesure où il constitue un paramètre à prendre en considération lors de la conception des tâches, comme des auteurs tels que Cuq (2003) et Reuter (2011) l'affirment. Par supports, nous désignons tous les documents oraux ou écrits, textuels ou iconiques que les enseignants sélectionnent pour une situation particulière. Les éléments constitutifs des supports (textes, illustrations, schémas, tableaux...) ont deux fonctions essentielles. La première est de décrire le contexte dans lequel les tâches attendues des apprenants se concrétisent : ce contexte particulier garantit la production de textes cohérents et adaptés à la situation donnée. La seconde fonction est de fournir les informations sur lesquelles les apprenants s'appuient pour réaliser les tâches fournies.

---

<sup>77</sup> Cette grille que Gérard (2009) sera employée dans la partie pratique de notre thèse pour évaluer la validité des situations proposées par les enseignants pour évaluer les compétences de leurs apprenants. Nous l'investissons également pour valider les situations évaluatives que nous proposerons.

La sélection des supports n'est pas une tâche aisée aux enseignants qui sont contraints à choisir des supports authentiques, préconisés dans tout enseignement/apprentissage, pour transposer en classe le caractère authentique des situations sociales. La nature des supports sélectionnés ainsi que leur nombre dépendent aux paramètres de la famille de situations à laquelle la situation en question correspond. Dans ce contexte, nous rappelons que les paramètres que les enseignants vérifient pour la validation des situations retenus englobent des précisions sur les supports à proposer (longueur du texte, nombre...).

### 3.2.2.2. Les tâches

La tâche constitue le deuxième élément constitutif des situations. Pour la définir, nous nous référons à Brousseau (2004), qui contestant les définitions associant tâche à situation, considère la tâche comme un ensemble d'activités que les apprenants réalisent. Cette conception est adoptée par des auteurs à l'instar de Nuan (1989) -cité par Cuq (2003)- ainsi que dans des travaux comme le CECRL (2011) qui emploie les deux notions comme synonymes.

Ces définitions qui présentent la notion de tâche tantôt comme une situation, tantôt comme une activité sont rejetées par certaines études. Selon Nonnon, la distinction entre tâche et situation s'impose dans la mesure où *«la notion de tâche spécifie ce qui dans une situation donnée contribue à définir, à orienter les élèves en l'inscrivant dans une dynamique de travail à l'apprentissage»* (1998, p. 56). Autrement dit, la tâche est une partie de la situation qui indique aux apprenants le résultat auquel les activités et les procédures à mettre en œuvre-implicitement ou explicitement indiquées- aboutissent. Il en est de même pour la distinction entre activité et tâche que des auteurs tels que Bouchard (1985), Leplat et Hoc (1983)...défendent. Contrairement à la tâche liée au travail prescrit, l'activité est relative à ce qui est réellement accompli par les apprenants. Ce qui nous amène à admettre, à la suite de certains auteurs comme Leplat et Hoc (1983), que l'activité renvoie aux opérations et procédures mises en œuvre par les apprenants pour exécuter la tâche prescrite par leur enseignant (Murillo, 2009). De ce qui précède, nous définissons la tâche -en retenant les diverses conceptions listées ci-dessus ainsi que celle de Reuter (2011)- comme le travail que les apprenants doivent réaliser pour répondre aux attentes de leur enseignant. Ainsi, ils enchainent diverses activités pour atteindre un objectif précis.

Une définition satisfaisante de la notion de tâche ne pourrait négliger les quatre tâches listées par Leplat (2000) -reprises par Murillo (2009)- qui distingue :

- La tâche à réaliser,
- La tâche prescrite,

- La tâche redéfinie,
- La tâche effective.

Selon Murillo (2009), les deux premières tâches se rapportent à l'enseignant alors que les deux dernières sont liées aux apprenants. En effet, la **tâche à réaliser** -qui renvoie à ce qui doit être fait- et la **tâche prescrite** -qui désigne ce qui est demandé aux apprenants- sont conçues par le concepteur. La différence établie par l'auteur entre **tâche à réaliser** et **tâche prescrite** interpelle notre attention à ce niveau, d'où la nécessité de l'explicitier. En effet, la première renvoie à ce que le concepteur attend de ses apprenants alors que la seconde est liée à ce qu'il exige réellement en formulant la tâche attendue. Cette différence provient de l'écart entre ce qui est attendu et ce qui est exigé, un écart existant au niveau des deux tâches associées aux apprenants. Celles-ci comprennent **la tâche redéfinie**, qui désigne ce que les apprenants projettent de faire en s'appuyant sur leur représentation de la tâche prescrite, et la **tâche effective**, autrement-dit ce qu'ils accomplissent réellement. Ces tâches liées aux apprenants ne sont pas épargnées de l'écart évoqué supra étant donné que le travail réellement exécuté par les apprenants ne correspond pas toujours à ce qu'ils ont l'intention de réaliser. Mentionnons que la dernière catégorie de tâches listées par Leplat (2000) -ou tâche effective- est légitimement associée aux activités entreprises par les apprenants pour atteindre un but précis. Les écarts évoqués entre les deux catégories associées à chaque acteur sont abordés par Nonnon (1998) qui attire notre attention sur un autre écart entre la tâche prescrite par le concepteur et celle que les apprenants se construisent : la tâche redéfinie s'écarte parfois de ce que le concepteur prescrit pour multiples raisons (une formulation incorrecte, une mauvaise interprétation, un manque de concentration...).

Une lecture attentive des types de tâches nous permet de distinguer la tâche du concepteur de celle des apprenants. Certainement, en évoquant la conception des situations évaluatives, nous entendons particulièrement les tâches formulées par le concepteur. Nous ne nions pas le lien existant entre les deux dans la mesure où l'exécution de la tâche des apprenants dépend étroitement de celle du concepteur, qui pourrait recourir à deux niveaux de tâches dans les situations d'évaluation. Nous devons ces deux niveaux de tâches à Nonnon (1998) qui distingue :

- Les tâches du premier niveau : les consignes les expriment explicitement. Le but à atteindre, les procédures à emprunter et les acquis à mobiliser sont suffisamment clairs pour que les apprenants puissent se mettre au travail.
- Les tâches du second niveau : n'étant pas explicitement exprimées par les consignes, les contraintes, les procédures, les ressources...pourraient être précédemment



prévues par le concepteur ou émaner des échanges avec les apprenants, d'où leur complexité. Comparées aux tâches –assez simples- du premier niveau, les tâches du second niveau pourraient s'écarter de ce qui est réellement attendu de la part du concepteur (Nonnon, 1998).

Il semble que les deux niveaux de tâches, ci-dessus cités, varient au niveau de la prescription : le niveau de prescription dans les secondes tâches est inférieur à celui des premières. Leplat et Hoc (1983) associent ce niveau de prescription à la nature des apprenants. Ils estiment que la prescription serait plus lacunaire avec des apprenants expérimentés, contrairement aux apprenants novices, qui pour exécuter une même tâche, ont besoin de plus de précisions (Murillo, 2009). Précisions que la prescription, aussi détaillée soit-elle, ne pourrait en aucun cas intégrer tous les détails du travail. C'est ce que des auteurs comme Rabardel, Carlin, Chesnais, Lang et Pascal (1998) insinuent en affirmant l'écart inévitable entre le travail prescrit et le travail réel.

Pour que les apprenants soient capables de redéfinir la tâche prescrite par leur enseignant, le concepteur doit tenir compte de plusieurs paramètres au moment de sa conception. En y ajoutant le résultat de la tâche, Cuq les énumère dans la citation suivante : « *concevoir une tâche implique la prise en compte de six paramètres : les objectifs, le support, les activités, les rôles respectifs des enseignants et des apprenants et le dispositif* » (2003, p. 234). Pour le premier paramètre, il est connu que les tâches prescrites en évaluation ne doivent pas s'éloigner des **objectifs** fixés étant donné que toute évaluation porte sur ce qui est enseigné. Proposer des tâches appropriées à un public donné requiert, en effet, la prise en considération des objectifs de formation retenus. Quant aux **supports** sélectionnés, leur sélection obéit à de nombreuses conditions. Le CECRL (2011) énumère les facteurs à considérer avant de proposer un texte à un public particulier. Il s'agit de : l'intérêt des apprenants, le type du texte, sa longueur et sa complexité linguistique, la présentation matérielle... A ces facteurs, nous ajoutons la cohérence du support avec la tâche prescrite : le support doit fournir des données permettant aux apprenants de réaliser ce qui est demandé. En plus des deux paramètres expliqués, s'ajoutent les **activités** et le **dispositif** à mettre en application pour la résolution de la situation-problème proposée. Le concepteur pourra les formuler explicitement ou laisser aux apprenants la tâche de les identifier. Enfin, les **rôles** assignés aux deux acteurs impliqués sont clairement définis au préalable : la tâche est à exécuter par les apprenants qui pourraient à certains moments bénéficier des directives de leur enseignant. Pour le **résultat** de la tâche, il est clair que le concepteur le définit pour conduire les apprenants à enchaîner compétences,

activités et stratégies pour y parvenir. Mentionnons que nous ne disposons pas de références théoriques pour appuyer notre explication des paramètres de Cuq (2003), exposée ci-dessus.

Quelles tâches les enseignants choisissent-ils ? Les tâches ne sont pas toutes du même niveau de difficulté : certaines sont plus difficiles que d'autres. C'est pour cette raison qu'il est impératif de sélectionner des tâches dont le niveau de difficulté est adapté aux apprenants (Murillo, 2009). Proposer à des apprenants expérimentés des tâches complexes et aux apprenants novices des tâches abordables est certainement nécessaire pour un aperçu fidèle sur les compétences de ces évalués. Cette adaptation est nécessaire en évaluation. Cependant, il existe une autre adaptation qui fausse tout le processus évaluatif entrepris.

De nombreuses études se sont intéressées à cette adaptation des tâches choisies au niveau des apprenants pour orienter les résultats obtenus. Certains concepteurs ont tendance à choisir telles tâches ou questions pour obtenir tels résultats. Ainsi, Crahay (1996) cite Posthumus (1947) qui lie les tâches sélectionnées aux résultats visés : les tâches retenues visent une distribution gaussienne des notes. Connaissant leurs apprenants, ils leur soumettent des épreuves aboutissant à cette distribution. Selon De Landsheere (2000), l'enseignant choisit ses questions en évaluation en tenant compte des connaissances de ses apprenants : il ne pose pas des questions sur les éléments ignorés par tous ou connus par tous. Toutes ces études confirment une idée formulée par Murillo qui estime que « *les enseignants sont conduits à choisir des tâches en fonction des résultats attendus de leurs élèves (anticipation)* » (2009, p. 91). Cette tendance, actuellement existante, ne conduit pas à des résultats valides répondant aux attentes réelles des institutions sociales. La société est certainement plus soucieuse et avide de résultats réels que d'enseignants de bonne réputation. Il est nécessaire de mentionner que le niveau des apprenants ne constitue pas le seul facteur déterminant la sélection des tâches à retenir en évaluation. Certains éléments comme les caractéristiques de la tâche et le temps disponible sont, en effet, aussi importants que le niveau des apprenants (Murillo, 2009).

Pour conclure cet élément, nous rappelons que la distinction entre activité et tâche, mise en exergue par des études multiples, est dépassée dans la mesure où le terme tâche est, selon Nonnon (1998) désigne les activités d'application, et que des documents de grande ampleur continuent à employer invariablement les deux concepts.

### **3.2.2.3. Les consignes**

Le dernier élément dont une situation d'évaluation est constituée est la consigne. Celle-ci ne pourrait être définie séparément de la notion de tâche : elle en est l'expression linguistique. Dans ce contexte, Cuq définit la consigne comme un « *discours visant la réalisation d'une*

*tâche* » (2003, p. 53). C'est à travers la consigne que les apprenants comprennent la tâche à réaliser et les attentes de leur enseignant. D'ailleurs, une bonne réalisation de la tâche demandée dépend dans certains cas d'une bonne compréhension de la consigne d'où la nécessité de veiller à une formulation correcte, explicite et claire de la consigne. A ce niveau, nous mentionnons que la consigne ne se réduit pas à une question. En effet, elle apporte -selon Gérard (2009)- des précisions sur la production attendue, les procédures à mettre en œuvre, les conditions de réalisation...

Réalisant l'importance de la consigne dans l'évaluation des compétences des apprenants, de nombreux auteurs à l'instar de Gérard (2009) accordent un intérêt aux exigences de la formulation des consignes. Selon cet auteur, le concepteur doit être *clair, adapté, contraignant, complet, concis et modeste*<sup>78</sup>. De ces conditions énumérées, nous relevons trois caractéristiques essentielles pour que la consigne soit correctement formulée :

L'adaptation : cette première caractéristique n'est pas particulière à la consigne dans la mesure où les supports et les tâches sont également sélectionnés en tenant compte de deux facteurs essentiels : les particularités du groupe évalué et les conditions du déroulement. Pour le premier facteur, Gérard (2009) précise que la consigne tient compte du niveau des apprenants : une consigne proposée à des apprenants novices ne serait adaptée à des apprenants expérimentés et inversement. Cette adaptation au niveau est perceptible dans la formulation linguistique de la consigne : le vocabulaire investi ne serait pas aussi complexe avec des publics variés. Le second facteur est aussi important que le premier dans la mesure où la consigne doit être adaptée à la contrainte temporelle du déroulement de l'évaluation. Nous savons depuis Murillo (2009) que l'enseignant-concepteur, connaissant ses apprenants, pourrait prévoir la durée nécessaire à la résolution d'une tâche donnée. Ce respect de la durée doit être respecté au niveau des consignes : il doit également veiller à ne pas proposer une consigne dont la résolution requiert plus que la durée fixée. Tous ces deux facteurs pourraient certainement être enrichis par d'autres assurant la formulation d'une consigne adaptée.

La clarté : cette deuxième caractéristique requiert une formulation directe et précise de la consigne. La formulation directe et claire est une condition primordiale étant donné que la compréhension de la consigne ne devrait pas constituer un obstacle à la résolution de la situation-problème. Le concepteur veille à ce qu'il n'y ait pas de pièges empêchant la compréhension de ce qui est attendu. C'est pour cette raison qu'il est préconisé d'éviter les

---

<sup>78</sup>Tous ces éléments requis lors de la formulation des consignes sont expliqués par Gérard (2009).

questions indirectes induisant les apprenants en erreur ou les questions ambiguës susceptibles de plusieurs interprétations. La clarté de la consigne est également conditionnée par sa complétude : le concepteur doit éviter une consigne lacunaire, donc imprécise, car les apprenants sont contraints à deviner des éléments qui pourraient les détourner du résultat réel de la tâche retenue. Pour ce faire, il est préférable de recourir à des consignes précises anticipant et répondant à toutes les interrogations éventuelles des évalués. Au niveau de cette deuxième caractéristique également, le poids du vocabulaire est à prendre en considération dans la mesure où certains mots mal choisis pourraient générer des incompréhensions ou des interprétations erronées des consignes soumises aux apprenants. Pour s'assurer de la clarté des consignes, le concepteur pourra recourir à une lecture critique réalisée par ses collègues pour les reformuler en cas d'imprécisions constatées. Dans certains cas, les lacunes ne sont relevées que par les apprenants lorsque le concepteur leur propose la situation d'évaluation. A ce niveau-ci, une reformulation plus claire est souvent tolérée pour que les apprenants manifestent leurs compétences réelles.

La concision : cette dernière caractéristique ne contredit pas les précédentes dans la mesure où le concepteur pourra proposer une consigne claire et concise. Autrement dit, une consigne qui précise ce qui est attendu des apprenants, en évitant tous les éléments, expressions ou détails superflus. Encombrer une consigne de ce qui est superficiel la rend vague et peu précise.

Ces trois caractéristiques de la consigne, expliquées supra, sont extraites des conditions listées par Gérard (2009) qui ajoute une condition extrêmement importante dans le cadre d'une évaluation de compétences : c'est la **règle des 2/3**, qui sera succinctement expliquée dans les paragraphes qui suivent.

En évaluation par situations complexes, le jugement de maîtrise d'une compétence donnée est émis en tenant compte de la maîtrise des critères définis pour évaluer les productions des apprenants. Cependant, le critère n'est maîtrisé que si les apprenants le réussissent au moins deux fois sur trois occasions offertes. C'est la règle des 2/3 intégrée par De Ketele (1996), Gérard (2009), Roegiers (2005)... Cette règle consiste à présenter aux apprenants dans une situation évaluative complexe trois occasions pour la vérification de la maîtrise des critères définis, et de cette manière la maîtrise de la compétence. Cependant, une condition est indispensable pour une évaluation rigoureuse de la maîtrise du critère : les occasions doivent être *indépendantes*. Selon Gérard (2009), cette indépendance requiert des occasions distinctes ou dissociées. Autrement dit, la réussite d'une occasion ne doit pas dépendre de celle de la

situation précédente. A ce niveau, l'auteur évoque les questions à tiroirs dans lesquelles chaque nouvelle question dépend de la réponse donnée à celle qui la précède.

Cette règle, à première vue difficile à appliquer dans une épreuve, pourrait être remplie si le concepteur opte pour trois tâches (ou trois consignes) distinctes dans la même situation. De cette manière, chaque tâche aboutit à une production permettant la vérification des critères définis. Mais, pour quelles raisons la vérification d'un même critère à plusieurs reprises est-elle requise ?

Pour répondre à cette question, nous présentons les limites du recours à une seule occasion pour vérifier la maîtrise du critère. Il est connu que réussir un critère donné dans une occasion ne reflète pas nécessairement sa maîtrise dans la mesure où cette réussite pourrait être due à des facteurs autres que la compétence (Rey, 2012). La généralisation et le jugement de compétence suite au constat de cette seule réussite ne seront pas fondés. Il en est de même dans les cas d'échec : un apprenant qui échoue dans une occasion n'est pas forcément incompetent étant donné que cet échec pourrait être généré par plusieurs facteurs tels que la fatigue, la mauvaise formulation de la consigne... Ainsi, une généralisation fondée sur un seul constat devient indubitablement abusive.

Conscient de ces limites, De Ketele (1996) propose la règle des 2/3, qui a un avantage incontestable : elle concilie le droit à l'erreur et la preuve de compétence. Cette règle ne bannit pas l'erreur qui est inévitable, procurant ainsi un climat de sécurité affective aux évalués. En même temps, elle préserve la répétition de la performance pour un jugement de compétence étant donné que la répétition de la performance est une preuve de la compétence des apprenants.

Nous obtenons ainsi les conditions nécessaires dont il faut absolument tenir compte lors de la rédaction des consignes d'évaluation. Avant de la soumettre aux évalués, il est nécessaire de procéder à une évaluation de cette consigne (par des collègues, des apprenants distincts de ceux qui passent l'examen...). Dans le cas échéant, le concepteur pourra concevoir une grille évaluant les situations d'évaluation ainsi que les consignes intégrées.

### **3.3. Conception de situations évaluatives : quelles étapes ?**

Dans ce qui précède, les composantes des situations évaluatives ainsi que les paramètres communs entre les situations issues d'une même famille sont précisées. Cependant, le concepteur ne pourrait procéder à cette conception des situations et à leur validation avant de définir le *référentiel de situations*, c'est-à-dire le profil de sortie. Les étapes suivantes de la conception d'une épreuve d'évaluation, constituée d'une ou de plusieurs situations, sont empruntées à Gérard (2009).

### **3.3.1. Identifier la compétence évaluée et les caractéristiques mobilisées**

Dans le milieu scolaire, le référentiel de compétences doit être élaboré en s'appuyant sur le *référentiel de situations*. Celui-ci renvoie aux situations de la vie scolaire et sociale dans lesquelles les apprenants auront à agir en FLE. Une fois élaboré, ce référentiel de situations est traduit en référentiel de compétences couvrant toutes les compétences que les apprenants doivent maîtriser pour agir efficacement.

Pour chaque public, des compétences précises sont développées puis évaluées. Chaque compétence est enseignée dans le cadre d'une situation puis évaluée en tenant compte des paramètres de la famille de situations dont ladite situation est issue. La première phase dans la conception d'une épreuve évaluative (constitué d'une ou de plusieurs situations) consiste à déterminer la compétence à évaluer (décrire un lieu, justifier un point de vue, expliquer un phénomène...). Il convient également de déterminer si les ressources que la situation mobilise sont déjà maîtrisées par les apprenants. En identifiant la compétence à évaluer et les ressources à mobiliser, le concepteur pourra procéder à la conception d'une situation évaluative en tenant compte des paramètres nécessaires. Rappelons que l'enseignant pourra sélectionner une seule situation avec plusieurs tâches, ou plusieurs situations proposant chacune une tâche précise. Dans les deux cas, une condition primordiale ne doit pas être négligée : la règle des 2/3.

### **3.3.2. Élaborer une situation conforme aux exigences**

La compétence à évaluer identifiée, le concepteur entamera la conception de la situation ou des situations. Les éléments qui la constituent sont conçus en tenant compte des paramètres de la famille de situations ainsi que des ressources à mobiliser.

Les supports : les programmes, privilégiant une entrée par les types, aboutissent à la production d'énoncés en FLE, relevant de types textuels précis. Pour exécuter cette tâche, les enseignants recourent à des documents variables (des illustrations, des schémas, des tableaux, des textes...), contenant des informations sur le thème traité. S'agissant des textes sélectionnés, ils ne relèvent pas forcément du type textuel retenu dans la mesure où ce qui importe le plus est l'utilité des informations contenues. En effet, les apprenants ont besoin d'informations pour les exploiter lors de la réalisation de ce qui leur est demandé. Le contenu informatif des textes ne constitue pas l'unique critère à prendre en considération dans la mesure où le concepteur ne néglige pas la longueur du texte, les caractéristiques du public, la contrainte temporelle... Chacun de ces éléments a un impact sur le rendement des apprenants qui réalisent les tâches demandées. Ainsi, le nombre des textes et leur longueur, s'ils ne sont pas adaptés au niveau des apprenants et à leur âge devient un obstacle qui les décourage...

Les tâches et les consignes : indépendantes, ces tâches doivent aboutir à des productions assurant une vérification de la maîtrise des critères déterminés. Les caractéristiques de ces tâches sont évoquées par Leroux et Bigras (2003), qui ont regroupé les caractéristiques listées par de nombreux auteurs (Louis, 1999 ; Wiggins, 1993 ; Popham, 1995...). Ainsi, une tâche doit être contextualisée, c'est-à-dire ancrée dans un environnement défini et explicité aux apprenants. Le contexte choisi est souvent significatif pour eux, les poussant à s'impliquer dans la tâche proposée. Celle-ci doit également représenter un défi, présentant un ou plusieurs obstacles à surmonter. Ces obstacles, contrairement à ce que certains peuvent croire, sont loin de décourager les apprenants : ils suscitent chez eux plus de motivation en vue de les surmonter. C'est pour cette raison que des auteurs comme Beckers (2002), Laurier, Tousignant et Morissette (2005) intègrent cette deuxième caractéristique. Rappelons que les tâches proposées ne doivent pas être compliquées ; mais plutôt complexes, mobilisant les opérations cognitives de haut niveau (Beckers, 2002). De ce point de vue, nous ne pouvons admettre comme tâches les activités d'application vérifiant le savoir-faire des apprenants, bien que l'usage commun se serve du terme *tâches* pour les désigner. A ces caractéristiques, s'ajoute l'authenticité des tâches qui suppose un lien étroit entre ce qui est proposé dans l'univers scolaire et ce qui existe dans le milieu social. Que pouvons-nous constater au niveau des caractéristiques des tâches évoquées supra ? Il est apparent que les caractéristiques des tâches présentent des ressemblances avec celles des situations évaluatives. Ceci est évident étant donné que les situations évaluatives, pour être authentiques, motivantes, signifiantes..., doivent être constituées de tâches et de supports qui le sont également.

Lors de la conception des situations, il ne faut pas négliger les paramètres de la famille de situations qui doivent être respectés. Parmi ces paramètres, nous distinguons les critères de correction qui doivent être définis.

### **3.3.3. Déterminer les critères de correction**

Suite à la conception de la situation évaluative, les concepteurs doivent déterminer les critères de correction qui doivent être appliqués à toutes les situations correspondant à une même famille. Ainsi, en évaluation des langues, Roegiers (2005) énumère quelques critères souvent intégrés dans la correction des productions : la pertinence, la cohérence, la correction de la langue... Ces critères se multiplient en indicateurs dont la précision est indispensable pour vérifier la maîtrise des critères. A ce niveau, nous ne nous étalons pas sur la distinction entre critère et indicateurs, les conditions de leur détermination et sur les modalités de correction des

situations évaluatives ; éléments que nous traiterons de manière détaillée dans la suite de notre travail.

### **3.3.4. Vérifier si la situation correspond aux exigences...**

Une évaluation de la situation s'impose avant de la soumettre aux apprenants évalués. En effet, il est nécessaire de s'assurer que cette situation mobilise les ressources préalablement apprises, qu'elle ait toutes les caractéristiques requises dans une situation et qu'elle corresponde à la famille de situation dont elle est issue. Cette validation s'opère de manières variées : Gérard (2009) propose une grille descriptive, dont les enseignants peuvent se servir pour évaluer leurs situations évaluatives. Cette grille, dont nous nous servons dans notre partie pratique, ne constitue pas le seul moyen dans la mesure où d'autres procédures aussi efficaces sont proposées par Gérard et De Ketele (2005). En effet, il est possible de recourir aux types de validation (la validation a priori, la validation empirique interne et la validation empirique externe), ou à des procédures d'expérimentation pour parvenir à une version finale exempte d'incorrections sur tous les niveaux.

### **3.3.5. Expérimentation**

Bien que les épreuves soient souvent élaborées en impliquant tous les enseignants chargés de la matière évaluée, elles ne peuvent être déclarées valides si elles ne sont pas expérimentées. Plusieurs types d'expérimentation sont distingués. Le premier type, le test pilote, est évoqué par Gérard (2009). Il explique que cette procédure, *légère* selon son expression, pourrait être investie par les enseignants pour la validation des épreuves et des situations qui les constituent. Proposée également dans le « *Manuel pour l'élaboration et la passation des tests et d'examens de langues* », elle consiste à soumettre la première version du test à un groupe restreint (de 4 à 5 personnes) pour relever certaines lacunes. Le deuxième type évoqué dans ce document est la *consultation des collègues*. Ce type-ci est proche de la *validation a priori*, proposée par Gérard et De Ketele (2005) étant donné qu'il consiste à soumettre la première version à des collègues qui donnent leurs remarques et leurs commentaires. Le troisième type est le prétest. Celui-ci consiste à proposer la version du test (la situation conçue) à un public qui présente les mêmes caractéristiques (âge, niveau ...) que le public concerné par l'examen. En plus des caractéristiques identiques des publics, les enseignants veillent à la passation du test dans les mêmes conditions pour éliminer tous les facteurs qui biaisent les résultats de l'expérimentation. Les interrogations des apprenants, leurs commentaires, leurs demandes d'explication pendant la passation ; les réponses aux questions posées, les lacunes constatées, les erreurs récurrentes... fournissent aux concepteurs des



renseignements utiles. Grâce à cette expérimentation, ils sauront si la situation est adaptée au niveau des apprenants évalués, si les supports fournissent les informations dont les apprenants ont besoin, si les ressources que la situation mobilise se limitent à celles préalablement apprises, si les consignes sont clairement formulées, si les tâches évaluent les compétences enseignées en classe, si le temps imparti à l'épreuve est suffisant... Toutes ces informations sur la situation proposée à un public –dont les caractéristiques sont similaires à celles du public auquel ce test est destiné- sont utiles et contribuent considérablement à son amélioration.

Cette expérimentation aboutit à la version finale du test qui prend en considération les propositions, les commentaires... Quoiqu'elle paraisse théoriquement facile à mener, nous devons reconnaître la difficulté de sélectionner un public-témoin présentant des caractéristiques identiques à celle du groupe concerné par l'examen vu l'hétérogénéité des apprenants.

En suivant ces étapes que nous citons en nous référant à Gérard (2009), les enseignants conçoivent des situations évaluatives conformes aux exigences assurant une évaluation valide de la compétence déterminée, tout en tenant compte des objectifs institutionnels et des caractéristiques des apprenants.

#### **4. Passation des épreuves**

Si nous intégrons cette phase dans les étapes de la conception d'un dispositif d'évaluation, c'est pour son rôle incontournable dans l'équité et la validité des résultats auxquels les épreuves conçues aboutissent. En effet, les épreuves doivent se dérouler dans des conditions propices permettant à chaque apprenant de manifester sa compétence, en dehors de toute contrainte. Les conditions de passation, dont il est question, concernent tous les éléments du déroulement : la présentation matérielle des épreuves, les conditions de passation, les instructions...

S'agissant de la présentation matérielle, il faut veiller à une bonne présentation des sujets d'examen (lisibilité de l'écriture, clarté des illustrations, les couleurs...). Si les concepteurs optent pour une présentation numérique le jour de l'examen, celle-ci doit être aussi claire et pertinente que la précédente. D'autres conditions de passation sont à prévoir pour assurer le bon déroulement du test. Parmi ces conditions, citons la nécessité de solliciter un nombre suffisant de surveillants et de correcteurs. Il en est de même pour le lieu et la durée de l'épreuve qui doivent être appropriés. Ainsi, la durée prévue doit être appropriée au niveau de difficulté de l'épreuve ou au nombre de tâches ou de situations à proposer. Pour les lieux de passation, il faut les aménager pour que toutes les conditions soient remplies assurant par-là un rendement effectif des apprenants, révélant leurs compétences. Enfin, il est nécessaire de bien expliquer

toutes les instructions (matériel autorisé, consignes...) avant le début de l'examen pour ne pas interrompre le déroulement de l'examen par des recommandations ou des interrogations.

Assurer de bonnes conditions de passation est un moyen pour s'assurer que les réponses fournies par les apprenants interprètent leurs compétences réelles. Cependant, il est indéniable que certains facteurs, notamment d'ordre psychologique, affectent les performances de certains apprenants qui seraient inférieures à leurs compétences réelles.

## **5. Jugement : quels outils faut-il intégrer ?**

Cette étape de la conception de dispositifs évaluatifs -développée par Roy (2017)- correspond à la sélection des outils permettant de porter un jugement sur les performances manifestées par les apprenants. De ces outils, nous distinguons suite aux travaux d'auteurs tels que Leroux et Mastracci (2015), Scallon (1988, 2004) et Tardif (2006) : la liste de vérification, le barème de notation et la grille d'évaluation. Certainement, ce dernier outil retient essentiellement notre attention car il est le plus utile et le plus efficace pour porter un jugement suffisamment objectif sur la compétence évaluée.

### **5.1. La liste de vérification**

Elle se présente sous forme d'une liste d'indicateurs dont l'évaluateur vérifie la présence ou l'absence dans la performance manifestée par les apprenants. Nous préférons à ce niveau le concept indicateur à celui de critère dont se servent des auteurs comme Scallon (1988), Côté, Tardif et Munn (2011)... car les éléments attendus et précisés dans la grille doivent être directement observables dans le produit évalué, d'où la précision des énoncés qui expriment ce qui est attendu des apprenants. Dans ce contexte, Scallon (1988) précise que ces énoncés qui constituent la liste de vérification doivent être clairement formulés pour écarter toute interprétation secondaire.

Cet instrument est facile à manipuler dans la mesure où le correcteur coche la case oui ou non, selon la présence ou l'absence des éléments constitutifs de la liste. Dans cet ordre d'idées, Côté, Tardif et Munn –en se référant à Bélair (1993), Durand et Chouinard (2006)- associent la liste de vérification à une « *grille d'évaluation dont l'échelle est dichotomique : oui ou non* » (2011, p. 29). A cette facilité de manipulation s'ajoute son utilité pour évaluer les progrès atteints par les apprenants dans une durée précise. Néanmoins, ces deux avantages que présente la liste de vérification demeurent insuffisants pour la déclarer efficace dans tout dispositif évaluatif étant donné qu'elle est plus appropriée à l'observation qu'à l'évaluation. Dans cet ordre d'idées, nous nous référons à Scallon qui affirme que « *la liste de vérification est plutôt un outil d'observation qu'un outil d'évaluation* » (1988, p. 214).

Ces éléments cités rendent cette liste de vérification inappropriée à une évaluation sommative des productions évaluées. Celle-ci paraît plus conforme à une évaluation formative visant la vérification des progrès réalisés par les apprenants et le degré de maîtrise des indicateurs retenus.

## 5.2. Le barème de notation

Selon Cuq (2003), le barème consiste en une série de critères auxquels un nombre de points est associé. En dépit de son importance majeure, il est souvent négligé par certains enseignants qui doivent s'y soumettre lors de la correction. En effet, il constitue « *à la fois un guide pour le professeur et une certaine garantie d'équité pour les élèves* » (Merle, 2012, p. 222). Le correcteur se réfère à ce barème qui indique les exigences de la correction : le nombre de points à attribuer à chaque réponse, les sanctions éventuelles si les réponses des apprenants s'écartent du corrigé type, la répartition des points les parties d'une réponse... En somme, ce barème appuie le correcteur dans sa tâche garantissant ainsi une évaluation équitable étant donné que les apprenants sont évalués conformément aux mêmes exigences. Pour ceux-ci, le barème est un outil qui justifie la note obtenue, d'où la nécessité de le communiquer dans les épreuves orales ou écrites. A l'équité de l'évaluation soulignée supra s'ajoute une deuxième qualité que le barème de notation assure. Il s'agit de la fidélité ou la stabilité de la note que cet outil assure, selon Cuq (2003), lorsque plusieurs corrections sont effectuées, étant donné que les diverses corrections s'appuient sur le même barème.

En dépit de son utilité, le barème de notation qui assure la fidélité et l'équité de l'évaluation ne constitue pas une réponse satisfaisante aux attentes réelles des enseignants qui ont besoin d'instrument plus détaillé, plus informatif et plus efficace : c'est la grille d'évaluation qui englobe le barème de notation, selon Springer et Huver (2011). En effet, le barème de notation, qui s'impose comme un excellent outil pour évaluer des épreuves constituées de questions évaluant les savoirs et les savoir-faire des apprenants, semble déficient pour une évaluation des productions obtenues à travers des situations complexes.

## 5.3. La grille d'évaluation

Tout comme les deux instruments de jugement précédemment évoqués, la grille d'évaluation est un outil qui s'appuie sur des critères précis. Néanmoins, la grille, « *avec ses critères et ses sous-critères accompagnés d'échelons, est beaucoup près de l'évaluation-jugement que ne l'est la liste de vérification* » (Scallon, 1988, p. 181). Elle est également plus détaillée que le barème dans la mesure où elle « *prévoit plusieurs critères pour chacun des*

*niveaux du barème* »<sup>79</sup> (2005, p. 232). Ces deux particularités en font un outil plus utile et plus efficace pour évaluer les productions complexes des apprenants. Réalisant son importance, nous développons particulièrement cet instrument qui est plus adapté à une évaluation des productions réalisées par les apprenants, en réponse aux situations évaluatives complexes proposées. Dans les éléments qui suivent, nous définissons la grille d'évaluation, nous explicitons les raisons de son exploitation et nous précisons les étapes de sa conception en nous appuyant sur la littérature scientifique collectée.

### 5.3.1. Qu'est-ce qu'une grille ?

Plus adaptée à l'évaluation des tâches complexes, la grille critériée est conçue par Bourguet, qui se réfère à Scallon (2004), comme un outil qui « *permet de porter un jugement sur la qualité d'une production ou d'un produit, l'accomplissement d'une prestation ou d'un processus qui ne peuvent être jugés tout simplement bons ou mauvais comme dans le cas d'une question à correction objective* » (2017, p. 92)<sup>80</sup>. De ce point de vue, la grille devient l'outil le plus efficace à investir dans une évaluation de compétences qui ne pourrait être effectuée sans recourir à des situations et à des tâches complexes. Ces tâches-ci aboutissent à des productions qui ne pourraient pas être jugées objectivement en recourant aux barèmes de notation. Contrairement aux réponses que les apprenants formulent pour répondre aux questions évaluant les savoirs et les savoir-faire, jugées correctes ou fausses aisément, les énoncés produits dans les situations évaluatives, complexes et mobilisant des ressources variées, ne pourraient être évaluées objectivement sans recourir à des critères et des indicateurs précis diminuant la subjectivité.

L'efficacité de la grille provient essentiellement de son caractère détaillé : elle est répartie en critères qui se subdivisent en indicateurs, ou éléments observables, accompagnés d'échelle. Ces critères, indicateurs et échelle, dont la grille est constituée, contribuent à une évaluation qualitative et quantitative de la performance manifestée par les apprenants. Dans cet ordre d'idées, le Carrefour de la réussite au collégial (2004) évoque l'utilité de la grille qui décrit les qualités de la performance attendue, aide les apprenants dans la résolution des tâches, guide les enseignants lors de la correction et fournit des données plus précises sur la performance évaluée.

---

<sup>79</sup> La citation est extraite du document intitulé « *Des outils pour favoriser les apprentissages. Ouvrage de référence pour les écoles de la maternelle à la 8<sup>ème</sup> année* », publié par le ministère de l'Éducation et de la Formation de Manitoba (Canada).

<sup>80</sup> La même citation figure dans le document intitulé « *Élaboration d'une grille d'évaluation. Atelier pédagogique à l'intention des enseignants universitaires* », rédigé par Côté, Tardif et Munn (2011, p. 8).

Mentionnons que cet organisme distingue deux concepts : la *grille de correction* et la *grille d'observation*. Sa définition de la grille de correction, qui ne diffère pas de celle que proposent Côté, Tardif et Munn (2011), insiste sur la fonction de jugement assurée par cet instrument de mesure en situation d'évaluation sommative. La grille d'observation est, quant à elle, utile dans une perspective formative car elle sert à collecter des données observables à maintes reprises. Cette fonction que le Carrefour de la réussite au collégial (2004) attribue à la grille d'observation nous rappelle la liste de vérification qui constitue un outil pour observer les progrès réalisés par les apprenants. Des deux grilles distinguées, il est question dans notre travail de la grille de correction dont les fins cadrent avec nos objectifs de recherche qui porte sur les outils de mesure et de jugement à investir en évaluation sommative des compétences. Soulignons que le concept de grille de correction et celui de grille d'évaluation s'emploient dans notre étude comme synonymes.

### **5.3.2. Pourquoi le recours aux grilles d'évaluation ?**

De la définition de la grille d'évaluation, nous constatons que les auteurs auxquels nous nous sommes référée sont unanimes quant à son efficacité et son utilité, notamment en contexte d'évaluation de compétences, d'où la préférence de cet outil sur la liste de vérification et le barème de notation. Pourquoi faut-il donc recourir à une grille d'évaluation en évaluation de compétences ?

En premier lieu, il est prouvé que la grille d'évaluation –contrairement aux autres instruments- assure l'évaluation des tâches complexes qui mobilisent souvent des opérations cognitives de haut niveau (analyse, synthèse et évaluation) et aboutissent à des productions complexes (Côté, Tardif & Munn, 2011). Afin d'évaluer ces productions, les listes de vérifications et les barèmes de notation se montrent insuffisants pour recueillir des informations pertinentes et satisfaisantes sur le niveau de maîtrise manifesté par les apprenants. Recourir aux grilles d'évaluation -qui englobent les critères, les indicateurs et les échelles- assure des notes fiables et valides, fondées sur des critères précis. A ce niveau, nous rappelons la nécessité de recourir à des outils garants de données approfondies sur les productions des apprenants pour que les jugements émis et les décisions prises soient justes et valides. Ce dernier point nous conduit à évoquer la deuxième raison qui rend l'exploitation des grilles d'évaluation indispensable. Celles-ci guident les enseignants lors de la correction assurant par là une certaine équité : les productions des apprenants sont en effet évaluées suivant les mêmes exigences. Elles contribuent également à la fidélité et à la stabilité des résultats obtenus : les notes

attribuées à une performance quelconque par plusieurs correcteurs sont souvent stables. Pour appuyer nos propos, nous citons Riba et Mègre qui affirment :

Les grilles ont pour objet de garantir la fidélité de la notation, destinée à réduire autant que possible la subjectivité de l'évaluation, et à assurer la stabilité de la notation pour une version d'épreuve et un niveau de compétences données :

- D'un correcteur à un autre ;
- D'un candidat à un autre ;
- D'un lien à un autre ;
- D'une date à une autre. (2014, p. 67).

Roegiers (2005) recense les avantages que présente le recours aux critères d'une manière particulière, et aux grilles critériées d'une manière générale, en évaluation de toutes les disciplines. Selon lui, une évaluation critériée assure des notes plus justes, qui limitent les réussites et les échecs abusifs, valorisent les aspects positifs dans les productions des apprenants et contribuent à identifier leurs lacunes, ce qui rend aisé tout processus de remédiation.

Pour les enseignants, reconnaître l'utilité et l'efficacité des grilles évaluatives en évaluation de compétences constitue un premier pas vers une évaluation et un jugement valides. Cependant, le plus important est de concevoir des grilles qui répondent aux normes requises, d'où la nécessité de préciser les étapes à prendre en considération lors de cette conception, qui tient compte du but de l'évaluation. A ce niveau, nous citons Côté, Tardif et Munn (2011) qui estiment que les grilles conçues à des fins formatives sont souvent plus détaillées que celles élaborées en évaluation sommative.

### **5.3.3. Conception des grilles : quelles étapes faut-il suivre ?**

Afin de préciser les phases que tout enseignant doit suivre pour concevoir une grille d'évaluation, nous nous référons à Côté, Tardif et Munn (2011), qui listent cinq étapes que nous reprenons les éléments suivants. Selon elles,

La démarche d'élaboration de la grille d'évaluation comporte globalement cinq principales étapes : le choix des critères d'évaluation, le choix de l'échelle d'appréciation, la définition de la façon dont le jugement global sera portée, l'assemblage de la grille et son expérimentation (2011, p. 9).

Certains concepts que les étapes citées par les auteures incluent (critère, échelle...) méritent d'être définis. C'est ce que nous faisons, mais en intégrant leurs définitions dans les éléments qui suivent.

### 5.3.3.1. Le choix des critères d'évaluation

En vue d'assurer l'objectivité, la fidélité et l'équité de l'évaluation des épreuves, le recours à la grille critériée s'impose. Cette grille est constituée de critères dont le choix est le premier pas dans son élaboration. Sélectionner les critères est nécessaire car ceux-ci varient en fonction de la discipline, de la compétence à évaluer, de la tâche, du niveau des apprenants... comme l'affirment Riba et Mègre (2014). Ces critères peuvent cependant se ressembler dans certaines disciplines comme les langues qui obéissent à une même logique en évaluation, selon Roegiers (2005), ou au niveau de certaines compétences telles que la compréhension. Avant de préciser les critères que nous retenons en FLE, il est nécessaire de définir en premier lieu ce que nous entendons par critère. Pour ce faire, nous citons la définition de Gérard qui avance : « *un critère est un regard que l'on porte sur l'objet évalué, un point de vue auquel on se place pour évaluer l'objet. Le critère correspond à une qualité de cet objet* » (2009, p. 70). Il correspond en effet à la qualité attendue dans la production des apprenants, dont le correcteur vérifie la maîtrise. C'est la maîtrise des critères exprimant les attentes du correcteur qui détermine celle de la compétence évaluée : si les critères sont réussis, un jugement de compétence sera émis. Comment la maîtrise du critère est-elle déterminée ?

Pour répondre à cette question, certains auteurs tels que Côté, Tardif et Munn (2011) se servent du concept *éléments observables* alors que d'autres comme Scallon (1988, 2004), Gérard (2009), De Ketele (2005) et Roegiers (2005) lui préfèrent celui d'*indicateurs*. Ces indicateurs -qui correspondent aux manifestations concrètes et observables des critères dans les productions évaluées- apportent des informations sur la maîtrise des critères définis par le correcteur. Autrement dit, la maîtrise du critère est déterminée à travers celle des indicateurs qui le constituent. Dans ce contexte, Roegiers (2005) précise que le critère ne suffit pas pour évaluer la production des apprenants, d'où le recours aux indicateurs qui constituent des indices concrets sur la maîtrise du critère. Ainsi, toute grille comprend des critères qui se subdivisent en indicateurs.

Les critères que le correcteur choisit ont plusieurs caractéristiques que nous citons en nous appuyant sur les travaux de Gérard (2005, 2009). Selon cet auteur, les critères retenus doivent être pertinents (évaluant réellement la maîtrise de la compétence) ; indépendants (la maîtrise d'un critère ne dépend pas de celle qui le précède ou le suit) ; pondérés (d'une importance variée) et peu nombreux (pour assurer la faisabilité de l'évaluation). La sélection des critères qui constituent une grille, selon Côté, Tardif et Munn (2011), requiert en premier lieu la précision des éléments observables, ou indicateurs, caractérisant le produit attendu. Ce qui est attendu dans ce produit est néanmoins d'une importance variée, d'où la distinction

opérée entre deux catégories de critères : les critères minimaux et les critères de perfectionnement.

Pour élucider la différence entre ces deux catégories, nous nous référons aux définitions proposées par Roegiers qui avance : « *un critère minimal est un critère qui fait partie intégrante de la compétence, un critère requis pour déclarer l'élève compétent. Un critère de perfectionnement est un critère qui ne conditionne pas la maîtrise de la compétence* » (2005, p. 111). Le premier est ainsi indispensable pour juger compétent un apprenant, alors que le second est lié à une qualité du produit attendu qui n'est pas décisive pour la réussite. Autrement dit, la maîtrise de la compétence dépend de celle des critères minimaux : si ceux-ci sont maîtrisés, l'apprenant est déclaré compétent. La valeur de chacune de ces deux catégories de critères se concrétise dans leur nombre dans une grille et dans leur poids dans la correction et la notation.

S'agissant du nombre, de nombreux auteurs à l'instar de Roegiers (2005) et Gérard (2009) préconisent le recours à un nombre limité de critères qui assure une note assez juste. Roegiers (2004) justifie sa position par trois raisons essentielles. La première se rapporte à la correction : lorsque le nombre de critères est restreint, ceux-ci sont assurément intégrés dans la correction. Cependant, si les critères sont nombreux, les enseignants négligent certainement certains d'entre eux. La deuxième raison est liée aux effets de ces critères sur la qualité des productions attendues dans les situations d'apprentissage et d'évaluation. Ainsi, les apprenants sont souvent attentifs aux critères lors de la réalisation des productions demandées. Cependant, s'ils sont nombreux, ils ne peuvent jamais les prendre en considération. Enfin, la troisième raison citée par Roegiers (2004) est relative aux caractéristiques à respecter lors de la sélection des critères, parmi lesquelles leur indépendance. Cette caractéristique ne serait pas assurée si les critères sont nombreux car le correcteur risque de choisir des critères liés, nuisant ainsi à la validité de la correction. Toutes ces raisons conduisent au choix d'un nombre limité de critères. Gérard (2009) propose quatre critères : trois critères minimaux et un critère de perfectionnement. Etant donné que les critères minimaux sont plus décisifs pour émettre un jugement de compétence, la règle des  $\frac{3}{4}$  est proposée puis validée par des auteurs comme De Ketele (1996) et Roegiers (2000). Elle exige de ne pas accorder aux critères de perfectionnement plus que le quart du total des points (Roegiers, 2015) ; ce qui signifie que « *la note attribuée aux critères minimaux doit représenter au moins  $\frac{3}{4}$  des points de l'ensemble* » (Gérard, 2009, p. 100).

Nous mentionnons que la valeur d'un critère varie d'une discipline à une autre. Un critère minimal dans une discipline pourrait être de perfectionnement dans une autre. Il n'y pas seulement la valeur des critères qui varie : les critères eux-mêmes varient d'une discipline à une



autre. A cet égard, Gérard (2005, 2008, 2009) estime inutile de dresser une liste de critères valables pour toutes les disciplines étant donné que chaque discipline requiert des critères particuliers. Néanmoins, cet auteur ainsi que Roegiers (2004) sélectionnent des critères basiques applicables à toutes les disciplines : la **pertinence** (adaptation de la production à la situation) et la **correction** (conformité à la norme), qui auraient des appellations diverses, auxquels s'ajoutent la **cohérence** (absence de contradictions) et la **complétude** (intégration de toutes les parties nécessaires).

Contrairement à ces premiers critères, des critères comme **la qualité de la langue, la qualité de la présentation, l'originalité...** sont rangés dans la catégorie des critères de perfectionnement (Gérard, 2005). Rappelons que la distinction des critères minimaux de ceux de perfectionnement est étroitement liée à la discipline concernée. Quels critères (minimaux et de perfectionnement) faut-il retenir pour évaluer les compétences des apprenants en classe de FLE ? Cette question, nous y répondons dans notre volet pratique de notre étude où des grilles critériées seront proposées pour évaluer les productions des apprenants évalués.

### 5.3.3.2. La sélection de l'échelle d'appréciation

La deuxième étape dans la conception d'une grille critériée consiste à choisir une échelle d'appréciation. Celle-ci consiste selon Côté, Tardif et Munn en « *un continuum qui permet d'indiquer la qualité et la quantité des éléments décrits* » (2011, p. 17). Les éléments conventionnels qui constituent l'échelle, appelés **échelons**, renvoient à des degrés particuliers de maîtrise de la compétence évaluée. Le nombre des échelons varie d'une grille à une autre (de deux à six) selon les intentions du correcteur (le degré de précision voulu), les critères choisis... Cependant, nous pouvons affirmer en nous appuyant sur l'étude réalisées par les mêmes auteurs que si le nombre d'échelons augmente, la conception de la grille ainsi que son emploi deviendront difficiles.

Comme ces échelles décrivent quantitativement et qualitativement les compétences évaluées, une distinction est établie entre plusieurs types d'échelles, comme le représente le schéma suivant, emprunté des Côté, Tardif et Munn (2011), qui se sont inspirées des travaux de Durand et Chouinard (2006).

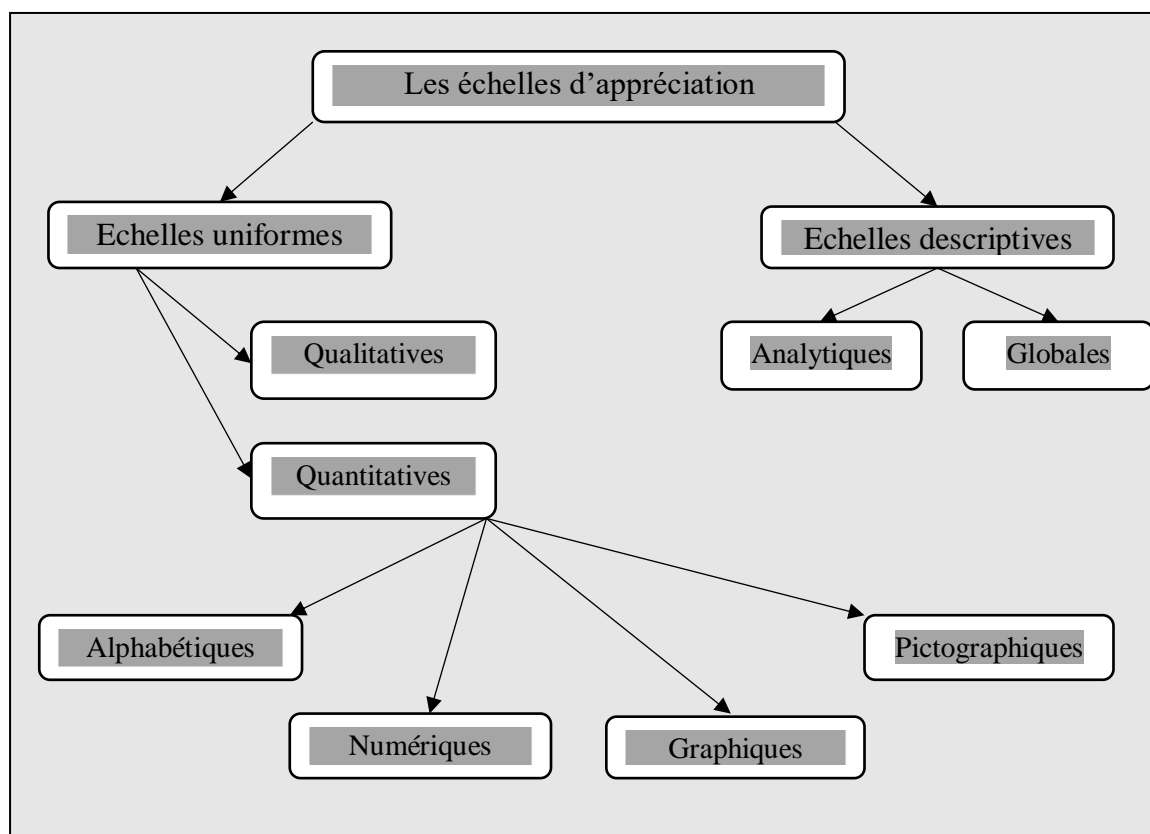


Figure 3: les types des échelles d'appréciation<sup>81</sup>

Les concepts désignant les divers types d'échelles d'appréciation que le schéma englobe sont expliqués dans les éléments suivants, en nous référant à Côté, Tardif et Munn (2011) :

- Les échelles uniformes

Très répandues, elles peuvent être qualitatives ou quantitatives. Les premières consistent à attribuer des appréciations qualitatives (médiocre, bien, très bien, excellent...) aux éléments observables, ou indicateurs. Ce sont des appréciations ou des qualificatifs dont le choix dépend de la nature des indicateurs retenus. Côté, Tardif et Munn (2011) dressent un tableau qui liste les qualificatifs que les enseignants pourraient associer à des indicateurs variés. Ainsi, des qualificatifs tels que *régulièrement, occasionnellement, jamais, toujours, rarement...* correspondent à la fréquence, alors que des appréciations comme *très au-delà des exigences, au-delà des exigences, conforme aux exigences, en deçà des exigences...* correspondent à un autre indicateur : le niveau de réussite<sup>82</sup>.

<sup>81</sup>Dans le schéma proposé par Côté, Tardif et Munn (2011), qui est extrait du travail de Durand et Chouinard (2006), la liste de vérification figure parmi les différents types d'échelles d'appréciation. Néanmoins, dans notre travail, la liste de vérification est conçue comme un instrument évaluatif à part entière, distinct de la grille d'évaluation, d'où son omission du schéma que nous reprenons.

<sup>82</sup> Nous ne reproduisons pas le tableau proposé par les auteurs Côté, Tardif et Munn (2011, p. 24) car les lecteurs pourront le consulter dans l'étude de ces auteures.

Fournissant des informations satisfaisantes sur chaque indicateur retenu, ces premières échelles sont les plus précises, selon Côté, Tardif et Munn (2011). En effet, elles décrivent la maîtrise de tous les éléments observables, ce qui informe les enseignants ainsi que leurs apprenants. Néanmoins, elles semblent inappropriées dans une perspective sommative d'évaluation étant donné que les détails fournis, qui ne sont pas concis, ne peuvent être aisément communiqués aux acteurs impliqués. Elles paraissent en effet plus adaptées aux intentions formatives des enseignants qui investissent les appréciations obtenues pour la régulation des apprentissages.

Les échelles quantitatives, quant à elles, indiquent avec précision le degré de maîtrise de ce qui est évalué. Elles se subdivisent en échelles numériques, alphabétiques, pictographiques et graphiques, dont les échelons sont représentés respectivement par des numéros, des lettres, des pictogrammes et des lignes. Cette sous-catégorisation est régie par la nature des échelons indiquant les degrés de maîtrise de la compétence évaluée.

Comme les échelles uniformes qualitatives, les échelles uniformes quantitatives sont largement expliquées par Côté, Tardif et Munn (2011) qui en fournissent des exemples précis. Ainsi, les échelles numériques et alphabétiques sont constituées de nombres (1, 2, 3, 4, 5 ou 0, 1, 2) ou de lettres (A, B, C, D, E) attribués à chacun des éléments observables constitutifs de la grille. La limite de ces échelles réside, selon les mêmes auteurs, dans l'ambiguïté de la signification des nombres et des lettres attribués aux indicateurs retenus. Cependant, les échelons représentés par nombres ou lettres renvoient dans les esprits des correcteurs et des apprenants, auxquels les échelles adoptées sont communiquées, à des appréciations précises : ce sont des appréciations représentées numériquement ou graphiquement.

En dépit de leur limite soulevée ci-dessus, les échelles alphabétiques et numériques demeurent plus précises que les échelles graphiques. Celles-ci représentent les appréciations des évaluateurs sur une ligne pour indiquer la maîtrise des indicateurs retenus. Dans les échelles pictographiques, cette maîtrise est indiquée par des pictogrammes divers (●, ●●, ●●● ; ☺...). Bien qu'elles soient significatives, ces échelles ne fournissent pas des renseignements précis sur la maîtrise des indicateurs. Selon Côté, Tardif et Munn (2011), ces deux sous-catégories d'échelles sont plus adaptées à l'auto-évaluation et à la co-évaluation. Elles sont peu compatibles avec une évaluation sommative dont les résultats, communiqués à tous les acteurs, doivent élucider le degré réel de maîtrise des compétences évaluées.

De toutes les échelles uniformes, les deux premières sont les plus courantes en évaluation sommative des apprentissages alors que les autres sont introduites en évaluation formative, en auto-évaluation et en co-évaluation pour remédier aux lacunes constatées étant

donné qu'elles interpellent l'attention sur les indicateurs et par-là les critères qui ne sont pas maîtrisés par les évalués.

- Les échelles descriptives

Contrairement aux échelles uniformes qui représentent le degré de maîtrise atteint par les apprenants par des appréciations qualitatives ou quantitatives (par numéros, lettres, lignes ou pictogrammes), les échelles descriptives décrivent de manière détaillée et précise ce qui est attendu des apprenants ayant atteint tel ou tel degré de maîtrise. Autrement dit, la série des qualités et des comportements attendus à chaque échelon sont détaillés. En dépit de la difficulté de leur conception, ces échelles assurent, selon Côté, Tardif et Munn (2011), la fidélité des résultats de l'évaluation. Elles comprennent : des échelles analytiques et des échelles globales.

Dans les premières, la description de chaque indicateur est formulée de manière précise et concise, en adoptant la même structure de phrases, alors que dans les secondes, la description est fournie globalement par niveaux de maîtrise progressifs, auxquels les performances évaluées sont comparées. Contrairement aux échelles analytiques qui décrivent ce qui est attendu dans chaque échelon pour qu'un élément observable soit maîtrisé, les échelles globales communiquent les exigences générales pour un jugement global de la performance évaluée. Des exemples pertinents présentant les particularités de ces échelles sont proposées par Côté, Tardif et Munn (2011).

Le caractère détaillé des grilles analytiques les rend utiles dans une perspective formative. En effet, elles permettent de détecter ce qui est maîtrisé ou peu maîtrisé, et ce qui ne l'est pas, fournissant des pistes de régulation. Contrairement à ces grilles, les grilles globales sont plus efficaces dans une logique de jugement car elles sont aisément communiquées aux acteurs intéressés.

En somme, la multitude des échelles d'appréciation complique aux enseignants la tâche de sélectionner celle qui répond le mieux à leurs objectifs. En effet, certaines échelles semblent plus utiles en évaluation formative car, par leur aspect formatif, elles fournissent aux apprenants des informations sur leurs progrès. C'est le cas des échelles uniformes qualitatives et des échelles descriptives analytiques. D'autres semblent, néanmoins, plus propices à intégrer dans une perspective sommative et certificative. C'est le cas des échelles uniformes numériques et alphabétiques très appliquées dans de nombreux systèmes éducatifs.

### **5.3.3.3. Déterminer un moyen pour exprimer le jugement**

Cette troisième étape dans la conception d'une grille d'évaluation est aussi importante que les précédentes. Suite à la sélection d'une échelle d'appréciation (les échelles uniformes

sont retenues), vient le moment de choisir le moyen d'exprimer le jugement global. Côté, Tardif et Munn (2011) évoquent deux moyens pour exprimer ce jugement : la note ou le score. Le choix du moyen adopté dépend de plusieurs circonstances telles que les habitudes évaluatives des enseignants et de leurs apprenants, les exigences institutionnelles... En effet, le correcteur ne pourrait en aucun cas adopter un moyen non-reconnu institutionnellement.

Dans le système éducatif algérien, l'évaluation sommative aboutit à une note qui représente le niveau atteint par les apprenants. Pour assurer la validité de cette note, qui présente un intérêt incontournable en évaluation sommative et certificative, le recours aux échelles numériques nous paraît un choix lucide. La note ne constitue pas une fin ; mais un moyen pour traduire le niveau de maîtrise atteint chaque apprenant. En effet, ce qui importe dans une évaluation à travers la grille critériée est d'évaluer la maîtrise des critères de la compétence. C'est dans cette optique que Gérard estime la tâche du correcteur consiste à attribuer une valeur à la production des apprenants à travers des expressions phrastiques du type « *l'élève maîtrise la compétence car les critères minimaux sont maîtrisés* » (2009, p. 99), une valeur traduite ultérieurement par une note.

Il est nécessaire de préciser certaines règles pour porter un jugement. La première est de fixer le seuil de réussite à appliquer, qui est défini en référence à Legendre (2005) comme le « *niveau de qualité à partir duquel on considère une performance comme réussie* » (Côté, Tardif & Munn, 2011, p. 10). Le correcteur pourrait joindre des règles supplémentaires comme exiger une note importante pour réussir un critère jugé important. Ces règles doivent être explicitées dans la grille<sup>83</sup>.

#### **5.3.3.4. Assembler la grille**

Cette phase concerne notamment la présentation de la grille. Dans ce contexte, Côté, Tardif et Munn (2011) énumèrent une série de précautions auxquelles le concepteur doit faire attention. Selon elles, il est nécessaire de veiller sur une bonne présentation et organisation des critères retenus, des espaces dédiés aux commentaires, aux renseignements importants (le nom de l'apprenant et celui du correcteur, la date...)... Si la grille est nouvellement introduite, il faut prévoir un guide qui explicite comment procéder pour une exploitation réussie de la grille en classe, afin d'atteindre les objectifs visés. Ces précisions destinées aux enseignants pourraient être insérées dans des notes de bas de page contenant également une explication de certains concepts-clés insuffisamment maîtrisés.

---

<sup>83</sup> Côté, Tardif et Munn (2011) proposent des exemples de grilles contenant des exigences liées au seuil de réussite.

### 5.3.3.5. Expérimenter la grille conçue

Pour la validation de la grille conçue, une expérimentation de celle-ci auprès d'un groupe d'apprenants est nécessaire. En évaluant certaines productions par la grille conçue, certains éléments pourraient être modifiés suite aux commentaires des apprenants, à qui tous les critères intégrés en évaluation sont communiqués, et aux remarques des enseignants. Cette expérimentation aboutit à une éventuelle modification de la grille qui pourrait porter sur : le nombre des critères ou celui des échelons, la formulation des indicateurs, suppression des éléments répétés...

En somme, les étapes que nous venons de lister ne se suivent pas forcément dans un ordre linéaire étant donné que certains choix pourraient se faire simultanément comme la sélection des critères et indicateurs ainsi que celle de l'échelle d'appréciation. Cependant, il est intéressant de ne pas négliger certaines phases telles que l'expérimentation, à première vue supplémentaire, qui pourrait contribuer considérablement au perfectionnement de la grille proposée au départ. Avec la conception de la grille critériée, retenue en tant qu'outil de jugement, s'achève la quatrième étape, le jugement, dans la conception d'un dispositif d'évaluation que Roy (2017) propose pour que la dernière étape débute : celle de la décision.

## 6. La décision

En évaluation sommative, cette phase est la plus importante -selon Bélair (2014)- car elle a des retombées directes sur l'avenir des apprenants : échec ou réussite, orientation en cas de réussite... Elle s'appuie sur l'interprétation des résultats obtenus pour déterminer leur réussite ou leur échec. La validité de cette décision est conditionnée par celle de toutes les étapes constitutives du processus en question. Du moment que cette dernière étape est le moment de prendre des décisions, les résultats des apprenants ainsi que les décisions prises doivent être communiqués à tous les acteurs impliqués. Pour ce faire, le recours à des outils comme le bulletin scolaire est nécessaire pour une communication réussie qui informe les apprenants, leurs parents ainsi que les autres acteurs impliqués. Mentionnons que Roegiers (2010) préfère voir le bulletin scolaire comme un outil de communication, pas uniquement d'information, car il amorce un dialogue avec les parents sur les compétences de leurs enfants<sup>84</sup>.

Cet élément de notre travail omet volontairement certains détails liés aux décisions prises par les enseignants et aux outils communiquant ces décisions aux acteurs concernés.

---

<sup>84</sup>Plusieurs outils autres que les bulletins sont introduits dans certains systèmes éducatifs tels que les portfolios. Nous n'avons cité que les bulletins car ce sont les seuls outils que le système éducatif algérien intègre.

Cette omission s'explique par la particularité de notre contexte d'étude, le système éducatif algérien, où les décisions de réussite ou d'échec, ainsi que celles d'orientation ne sont pas prises en fonction des résultats obtenus dans une seule matière (le FLE). En effet, c'est la moyenne générale de toutes les matières enseignées puis évaluées qui conditionne les décisions à prendre.

### **Conclusion**

En vue d'une évaluation rigoureuse des compétences des apprenants, il est nécessaire de concevoir des dispositifs évaluatifs conformes aux exigences de l'APC, une approche actuellement appliquée dans de nombreux systèmes éducatifs. Cette conception passe par plusieurs étapes (la planification, la conception, la passation, le jugement et la décision) dont certaines semblent plus décisives que d'autres, au moins du point de vue des enseignants-concepteurs. Certaines de ces étapes, inspirées de Roy (2017), comme la passation et le jugement ne dépendent pas uniquement des enseignants étant donné qu'elles impliquent également le personnel administratif des établissements scolaires. D'autres étapes -comme la planification, la conception et le jugement- sont étroitement liées au corps enseignant qui doit planifier les épreuves et concevoir les outils de mesure et de correction appropriés.

Étant donné que notre travail porte sur la validité des évaluations sommatives du FLE dans le système éducatif algérien, évaluations qui sont conçues par les enseignants de la matière, nous avons consacré aux étapes étroitement liées aux enseignants plus de pages que celles accordées aux étapes dont les enseignants ne sont pas les seuls responsables.

La planification, première étape du processus, constitue le pilier qui soutient toute la construction : la validité de tout le processus en dépend. Les enseignants y précisent la nature des objets qu'ils doivent évaluer, les contenus à évaluer, les conditions des épreuves... Dans une logique de compétences, les enseignants doivent déterminer si les épreuves du FLE à concevoir évaluent les compétences ou les ressources, ou les deux. S'agissant des compétences, contrairement aux croyances qui s'opposent au caractère évaluable des compétences, celles-ci sont évaluées à travers des performances, qui sont les manifestations concrètes des compétences des apprenants. Cette évaluation de compétences n'exclut pas celle des ressources dans la mesure où des auteurs tels que Gérard (2009) n'objectent pas une évaluation de ressources même dans le cadre de l'APC. Cette précision des objets sur lesquels l'évaluation porte détermine les outils de mesure que les enseignants élaborent lors de la phase de conception. A ce niveau, les ressources sont évaluables à travers des activités évaluant les savoirs (restitution des règles grammaticales) et les savoir-faire (application des règles dans des contextes variés), alors que les compétences le sont à travers des situations. Ce sont des situations évaluatives aux

caractéristiques précises que ce chapitre a abordé de manière détaillée. Ces situations complexes, significatives, authentiques, motivantes et complètes doivent être validées avant leur passation. Pour ce faire, les enseignants recourent aux procédures de validation des épreuves que De Ketele et Gérard (2005) exposent, procédures adaptées à la validation des épreuves évaluant les compétences. Mais, ils doivent également tenir compte des paramètres des situations sélectionnées qui sont communs aux situations appartenant à une même famille. Qu'elles soient activités de restitution, d'application ou situations complexes d'évaluation, les enseignants doivent les valider de manières diverses avant leur passation. Ce moment-ci du processus évaluatif instaure des conditions propices permettant aux évalués de manifester leurs compétences : de bonnes conditions de passation visent à réduire le gouffre entre compétences réelles et compétences manifestées.

Ce sont les compétences manifestées par les apprenants, ou leurs productions, qui sont corrigées par les enseignants. Pour cette étape de jugement, ils ont besoin d'outils qui évaluent le plus fidèlement possible les acquis de leurs apprenants. Il existe plusieurs outils de jugement dont le choix dépend des outils de mesure investis. Ainsi, les barèmes de notation conviennent à la correction des activités évaluant les savoirs et les savoir-faire alors que les grilles critériées sont les plus conformes à une évaluation par des situations complexes. Contrairement à la liste de vérification qui est plus utile en évaluation formative des apprentissages, ces deux outils sont les plus courants en évaluation sommative, bien que certaines catégories de grilles soient plus utiles dans une perspective formative d'évaluation. Ces outils aboutissent à la prise de décisions, une phase que nous n'avons pas développée suffisamment dans notre travail. Il est clair que ce chapitre ne répond pas à toutes les interrogations que suscitent les éléments traités. Certains aspects, qui n'y figurent pas, sont intentionnellement occultés pour ne pas encombrer notre étude de détails que nous jugeons superflus alors que d'autres seront traités dans notre volet pratique.

En abordant les étapes constitutives du processus évaluatif, il paraît à certains moments que la conception des outils de mesure et de jugement est un fardeau sous lequel s'écrasent les enseignants en raison de la complexité de leur tâche. Ceux-ci, qui se retrouvent en effet face à des pratiques nouvelles et étrangères aux milieux dans lesquels ils exercent, sont contraints de tenir compte de toutes les exigences pour parvenir à une évaluation des compétences de leurs apprenants. Cet objectif n'est cependant pas si aisé à atteindre étant donné que l'APC, dite nouvelle approche, inclut des pratiques enseignantes et évaluatives qui ne sont pas encore réellement intériorisées par les enseignants qui doivent payer le prix onéreux du changement. Le premier pas vers ce changement serait la formation des enseignants, anciens et nouveaux,



aux pratiques, évaluatives notamment et enseignantes si nécessaire, que l'APC requiert pour un meilleur développement des compétences des apprenants algériens.

# **PARTIE PRATIQUE**

# **CHAPITRE 1 : CADRE METHODOLOGIQUE**

## Introduction

Après avoir défini les concepts-clés liés à notre étude et en dessiné les contours dans notre premier volet théorique, nous entamons à présent le second volet pratique que nous répartissons en trois chapitres.

Rappelons que notre recherche porte sur la validité des dispositifs évaluatifs introduits dans le cycle moyen du système éducatif algérien en vue d'évaluer les compétences des collégiens en FLE. C'est la raison pour laquelle nous avons subdivisé le volet théorique en trois chapitres. Le premier a porté sur l'évaluation d'une manière générale et sur l'évaluation sommative d'une manière particulière en vue d'en déterminer les acteurs, les fonctions, les modalités, les critères... C'est à travers ce premier chapitre que nous avons précisé que notre étude s'intéresse à l'évaluation des apprenants, ou plus précisément de leurs compétences en FLE, au terme d'un ensemble d'apprentissage. Comme cette évaluation devrait être régie par des critères qui en assurent la validité, nous avons étudié tous les critères qui devraient se réunir en évaluation. De tous les critères, celui de la validité est retenu comme un critère suprême regroupant tous les autres qui n'en constituent que des dimensions. Ce critère de validité qui a occupé le deuxième chapitre de notre volet théorique renvoie à la capacité des outils de mesure et de jugement à donner un aperçu assez réel sur les compétences des évalués, à travers une évaluation compatible aux objectifs fixés par les enseignants, qui sont en réalité ceux visés par l'institution. Cette notion de validité a conduit à celle de validation qui vise à travers des procédures variées à prouver la validité d'un dispositif évaluatif. Ces procédures constituent également une des phases de la conception des dispositifs évaluatifs qui ont été largement évoquées dans le dernier chapitre théorique.

Jugeons ces éléments abordés dans notre volet théorique suffisamment utiles pour notre recherche, nous passons au premier chapitre de notre partie pratique que nous consacrerons au cadre méthodologique de l'étude : nous expliciterons les démarches adoptées, les outils de collecte de données investis, le contexte dans lequel s'inscrit notre recherche, les corpus à analyser... Ces éléments doivent être nécessairement traités de manière détaillée car les autres chapitres pratiques en dépendent.

Le premier élément que nous évoquerons dans ce chapitre est la méthodologie, c'est-à-dire l'« ensemble des méthodes et des techniques qui orientent l'élaboration d'une recherche et qui guident la démarche scientifique » (Angers, 2015, p. 84). De toutes les méthodes, qui sont multiples, nous aborderons celles auxquelles nous recourons dans notre étude en tentant d'en justifier la sélection, étant donné que chaque méthode répond à des objectifs précis. Il est

important de préciser que notre étude, à l'instar de toute étude scientifique, ne s'appuie pas sur une méthode unique et alterne des démarches variées. Cette présentation sera précédée de quelques précisions sur les types de notre recherche, que nous avons tenté de déterminer en nous référant à l'ouvrage d'Angers (2015) intitulé « *Initiation pratique à la méthodologie des sciences humaines* ».

Suite à la précision des méthodes investies, nous étudierons le contexte dans lequel s'inscrit notre recherche. Étant évident que notre recherche s'inscrit dans un contexte national, nous tentons à travers cet élément d'apporter des précisions supplémentaires sur les évaluations sommatives introduites et sur les moments de leur introduction. Ces précisions sont nécessaires pour deux raisons essentielles. La première concerne la multitude des formes évaluatives et des pratiques introduites au cycle moyen pour évaluer les diverses connaissances et compétences des apprenants, ce qui requiert la détermination de celles qui relèvent de la forme sommative. Quant à la seconde raison, elle se rapporte à la variété des pratiques d'un pays à un autre, qui rend indispensable la consolidation des informations théoriques du premier volet, regroupant des données globales, par des données particulières ancrées dans le contexte national.

En plus des éléments cités ci-haut, ce chapitre décrira profondément les *outils de recherche*<sup>85</sup> investis pour recueillir des données utiles à notre étude. Dans notre étude, nous retenons en effet deux grilles : une grille d'analyse et une liste de vérification. Nous les avons élaborés à fins précises, que nous aborderons, en nous appuyant sur la matière scientifique dont nous disposons. Nous justifierons également la sélection de certains critères et indicateurs constituant les deux grilles.

Ces éléments détaillés ne sont pas exhaustifs dans la mesure où d'autres précisions seront intégrées dans ce premier chapitre. Nous en avons listé les plus importantes qui seront largement évoquées dans les éléments qui suivent. D'autres informations et précisions jugées utiles et que nous n'avons pas mentionnées ci-dessus seront abordées même succinctement pour fournir dans ce premier chapitre toutes les précisions jugées nécessaires.

## **1. Méthodologie**

Notre recherche, qui porte sur la validité de l'évaluation sommative du FLE au cycle moyen en Algérie, est une étude disciplinaire qui s'inscrit dans le cadre de la didactique des langues étrangères. Relevant plutôt d'une *recherche-appliquée*, qui s'oppose à la recherche

---

<sup>85</sup> Ce sont les instruments investis en vue de collecter des informations quantitatives et qualitatives sur l'objet d'étude tels que les grilles d'analyse, les questionnaires...

fondamentale<sup>86</sup>, elle est orientée vers « *l'action, la mise en œuvre, la concrétisation des savoirs théoriques* » (Richer, 2011, p. 49). Dans le cas de notre étude, nous nous appuyons sur la littérature scientifique traitant notre thème pour parvenir à des propositions satisfaisantes. En effet, notre intention ne consiste pas à élaborer des théories ou des connaissances nouvelles ; elle consiste essentiellement à appliquer les modèles, procédures et démarches dont nous disposons pour étudier la validité des pratiques évaluatives actuelles dans notre système éducatif et pour concevoir des dispositifs évaluatifs nouveaux qui seraient plus conformes aux exigences des approches actuelles.

Avant de nous étaler sur la démarche adoptée, nous tenons à préciser que notre recherche est à la fois *descriptive* et *explicative*. Si nous nous servons du premier qualificatif pour la décrire, c'est en raison des détails et précisions que nous fournissons sur les évaluations sommatives que nous étudions. Quant au second qualificatif, son usage s'impose étant donné que des explications doivent être formulées pour justifier le recours à certaines pratiques évaluatives. C'est en associant ces deux perspectives que des données assez profondes pourraient être recueillies.

Tout ce qui est abordé ci-haut concerne notamment la nature de notre recherche dans la mesure où nous en avons déterminé les types et ce, en fonction de critères précis. Ainsi, notre intention a rangé notre étude dans les *recherches-appliquées* alors que notre visée l'a rendue *descriptive et explicative*. Il convient donc de passer aux précisions sur les approches adoptées en vue de collecter les données nécessaires et les analyser. Quelle(s) approche(s) faut-il investir pour mieux répondre aux questions de notre étude ?

### **1.1. L'analyse qualitative et quantitative (approche mixte)**

En vue de collecter des données dont l'analyse répond aux questions retenues, nous optons pour une approche mixte alternant les démarches quantitative et qualitative. Soulignons que les approches quantitatives cherchent à « *assurer la validité et la généralité d'une affirmation donnée que l'on appellera dans ce contexte une hypothèse* » (Mongeau, 2008, p. 31) alors que les approches qualitatives visent la compréhension de ce qui est étudié en interprétant les éléments observables (Mongeau, 2008 ; Angers, 2015).

---

<sup>86</sup> La recherche fondamentale, contrairement à la recherche-application, s'appuie sur les connaissances théoriques existantes pour développer des savoirs théoriques utiles. Ce sont les savoirs élaborés grâce à cette recherche qui seront ultérieurement mis en œuvre dans la recherche-application (Richer, 2011 ; Angers, 2015).

Le choix de *l'approche qualitative* se justifie par la nature des documents sur lesquels porte notre analyse. Nous procédons, en effet, à une analyse du contenu manifeste des épreuves écrites proposées pour évaluer les apprenants en FLE au cycle moyen. Cette analyse assure la compréhension des pratiques évaluatives introduites par les enseignants pour vérifier ultérieurement leur conformité aux exigences de l'approche actuellement appliquée (APC), autrement dit leur validité. Cette approche-ci, appliquée dans le système éducatif algérien à l'enseignement/apprentissage de toutes les disciplines, requiert, comme nous l'avons précisé dans le troisième chapitre de notre volet théorique, des outils particuliers pour une évaluation valide. Il devient nécessaire de vérifier l'intégration de ces outils à travers une analyse des épreuves collectées. En vue d'étudier le contenu des épreuves qui constituent notre corpus, nous recourons à une grille d'analyse constituée d'indicateurs ou éléments observables qui nous permettent une description qualitative des évaluations sommatives analysées. Nous investissons également une liste de vérification constituée d'indicateurs en vue de vérifier la validité des situations évaluatives que les sujets comportent ou que nous proposons ultérieurement. Ces deux instruments de collecte, que nous développons de manière détaillée dans ce qui suit, serviront à fournir des données essentiellement qualitatives à notre recherche.

Néanmoins, cette approche qualitative apporte des précisions sans pour autant permettre la généralisation de nos résultats, d'où la nécessité d'intégrer l'approche quantitative. En effet, nous représentons les résultats collectés à travers la grille d'analyse appliquée aux sujets analysés par des taux pour parvenir à des résultats soutenant la validité des résultats obtenus. La quantification porte sur les échelons mesurant chaque indicateur constituant la grille que nous représentons par des taux reflétant les tendances dominantes.

Ces approches qualitative et quantitative qui constituent la méthode mixte adoptée apportent des précisions sur les pratiques évaluatives adoptées et notamment sur leur validité. Néanmoins, elles doivent être consolidées par une méthode aussi utile pour la formulation de résultats assez fondés. Il s'agit de la méthode comparative que nous introduisons à des fins précises.

## **1.2. La méthode comparative**

Pour appuyer la méthodologie mixte investie dans notre étude, nous jugeons pertinent de recourir à la démarche comparative. Celle-ci « *fait de la comparaison l'unique moyen d'analyser les données concrètes afin d'en dégager les éléments constants* » (Saucin, 2018, p. 111). Elle aboutit au repérage des convergences et des divergences qui ne constituent certes pas la fin de cette démarche comparative dont la visée est de fournir des informations profondes sur

le phénomène étudié. Nous employons ce terme de *méthode comparative* dans une acception qui ne s'écarte pas de celle que nous venons d'évoquer. En effet, nous entendons par cette méthode comparative dans la présente étude la comparaison entre des documents analysés. Cette comparaison, à travers l'identification des ressemblances et des différences, nous permet de déterminer à quel point les pratiques ont évolué dans des moments différents. Bien que notre recherche semble revêtir un *caractère diachronique* de ce point de vue, nous ne la considérons pas comme telle étant donné que la recherche diachronique, contrairement à la nôtre, retrace l'évolution du phénomène étudié dans le temps (Angers, 2015).

Dans le cas de notre étude, cette approche s'avère pertinente pour confronter deux dispositifs évaluatifs : le premier était introduit à l'ère de la pédagogie par objectifs (PPO), tandis que le second correspond aux pratiques actuelles qui relèvent de l'APC. Ainsi, nous comparons des sujets anciens antérieurs à l'adoption de l'APC à ceux qui lui sont postérieurs. A ce niveau, le corpus qui servira notre étude comparative est réduit aux épreuves de BEM<sup>87</sup>.

Le recours à cette démarche se justifie par la nécessité de consolider les résultats auxquels nous parvenons à travers l'étude quantitative et qualitative des documents constituant notre corpus. En leur ajoutant les résultats de cette démarche comparative, les conclusions que nous formulons seront irréfutables car elles seront fondées sur des bases suffisamment solides.

Ces méthodes que nous adoptons dans la présente recherche visent la collecte de données répondant à nos questions et ce, en investissant des instruments divers que nous exposons dans ce qui suit.

## 2. Les instruments de mesure

En vue d'analyser les épreuves écrites qui constituent notre corpus, nous avons investi une grille d'analyse apportant des données qualitatives et quantitatives sur les pratiques évaluatives des enseignants au cycle moyen. A cette première grille, s'ajoute une deuxième investie pour une visée différente. Celle-ci, contrairement à la première, est essentiellement une liste de vérification que nous soumettons aux experts qui valident les situations évaluatives conçues. Son application assure le repérage des failles constatées pour les corriger. Ce sont ces deux instruments que cet élément tente de décrire.

---

<sup>87</sup>**BEM** ou Brevet d'Enseignement Moyen : examen qui clôt le cycle moyen dans le système éducatif algérien. Le cycle moyen, une étape transitoire entre le primaire et le lycée, s'étale sur quatre années qui correspondent à trois paliers. Le premier palier correspond à la première année moyenne, le second englobe la deuxième et la troisième année moyenne alors que le dernier palier correspond à la quatrième année moyenne. Au bout de ces quatre ans, les apprenants doivent réussir les examens du BEM pour passer au lycée.



## 2.1. La grille d'analyse

Cette grille a l'objectif d'étudier la validité des épreuves écrites que les enseignants du cycle moyen élaborent en vue d'une évaluation sommative des apprenants en FLE. Nous l'avons subdivisée en cinq parties dont chacune apporte des informations précises. Ce sont : les précisions sur l'épreuve, les compétences évaluées, le(s) support(s), les outils de mesure et les outils de jugement. Les indicateurs qu'englobe chacune des parties correspondent aux éléments observables qui fondent la validité des épreuves évaluatives. En effet, ces indicateurs sont extraits des informations théoriques précédemment évoquées sur la validité et la conception de dispositifs évaluatifs valides. Nous en justifions la sélection dans la présentation ci-dessous qui expose les parties qui structurent notre grille.

En plus des indicateurs extraits du contenu théorique de notre étude, nous devons évoquer à ce niveau l'apport indéniable des travaux réalisés par des auteurs algériens dans le domaine de l'évaluation scolaire. Citons l'étude de Ferhani-Meghraoui intitulée « L'évaluation des épreuves du baccalauréat à l'aune du Référentiel général des programmes », qui vérifie la congruence entre les exigences du Référentiel et les épreuves de français proposées de 2004 à 2014. Cette étude exploite une grille contenant des renseignements sur l'épreuve analysée (le type, l'année, la filière et la durée) ; des précisions sur les supports textuels sélectionnés (le titre, l'auteur et la source) et sur les questions posées (le type et le niveau taxonomique). Cette grille investie par l'auteure dans son étude a inspiré de nombreux éléments de celle que nous avons conçue.

### 2.1.1. Les précisions sur l'épreuve

La première partie de notre grille fournit des précisions sur les épreuves analysées. Nous avons jugé important de retenir cinq précisions essentielles : le type de l'épreuve, le niveau des évalués, le trimestre, l'année et la durée.

- Le type de l'épreuve

Trois types d'évaluation sommative sont précisés : devoirs, compositions et examens de BEM. Rappelons que les premiers sont introduits pour évaluer une séquence d'apprentissages, les seconds sont intégrés pour évaluer les apprentissages dispensés durant un trimestre alors que les derniers interviennent pour évaluer un cycle d'apprentissages. Cette distinction est nécessaire, à nos yeux, car elle permet de vérifier si les pratiques et activités auxquelles les enseignants recourent pour évaluer leurs apprenants varient selon les types, autrement dit, selon le volume des contenus enseignés. En effet, il serait intéressant de déterminer si chacun des types cités requiert des pratiques évaluatives particulières.

- Le niveau des évalués

Nous retenons cet élément pour préciser le niveau des évalués concernés par les épreuves analysées. En effet, le cycle moyen s'étend sur quatre niveaux qui visent des objectifs précis et des compétences particulières. La congruence entre les objectifs du programme et les items évaluatifs, pour être vérifiée, exige la précision des objectifs particuliers à un niveau donné, d'où la nécessité de le préciser. Aussi, il est nécessaire de préciser si une évaluation des compétences est introduite avec tous les niveaux du cycle ou seulement avec des niveaux précis.

- Le trimestre

Une précision sur les épreuves analysées est ajoutée : il s'agit du *trimestre*. Bien qu'elle paraisse peu utile de prime abord, cette précision pourrait apporter des données utiles à notre étude à travers notamment la comparaison des épreuves des différents trimestres. Cette comparaison, qui repère les particularités des épreuves d'un trimestre donné, nous informe si chaque trimestre privilégie des modalités particulières d'évaluation. Elle révèle également si le troisième trimestre, qui se situe au terme de l'année scolaire, évalue les objectifs ou se limite à une évaluation des objectifs trimestriels.

- L'année scolaire

Une autre précision est retenue dans la première partie : *l'année* des épreuves. A travers cette information, nous voulons nous assurer du caractère récent des épreuves analysées. Étant donné que l'application de l'APC en Algérie, qui remonte à 2003, n'est pas récente, toutes les épreuves retenues supposent une conformité à ses exigences. Néanmoins, scruter une éventuelle évolution des pratiques évaluatives au niveau des épreuves analysées, dont les dates varient, est un élément non négligeable dans la mesure où il répond à certaines de nos questions de recherche. Si certaines pratiques évaluatives se révèlent incompatibles avec les exigences de l'approche appliquée, nous saurons - au moins- si elles évoluent vers une meilleure application.

- La durée

C'est la dernière précision que nous retenons dans notre grille. De toutes les précisions précédentes, celle de la durée est la plus évoquée dans les informations précédemment citées sur la validité. Si nous la retenons, c'est parce qu'elle nous servira par la suite à vérifier si le temps imparti aux épreuves convient au nombre des items sélectionnés.

Ce sont les précisions que contient la première partie de notre grille. Celle-ci est suivie d'autres parties qui concernent essentiellement le contenu des épreuves analysées. Nous exposons dans ce qui suit les indicateurs qui se rapportent aux compétences évaluées.

### 2.1.2. Les compétences évaluées

A travers cette deuxième partie, nous déterminerons les compétences évaluées dans les épreuves analysées ainsi que les outils introduits pour les évaluer. Pour ce faire, les indicateurs que nous listons dans les points suivants sont retenus :

- La précision des compétences évaluées

En langues, deux compétences sont développées chez les apprenants, tant à l'oral qu'à l'écrit. Il s'agit de la compréhension et la production dont la maîtrise assure des interactions langagières efficaces dans les situations scolaires et sociales. A travers cet indicateur, nous saurons si les épreuves écrites analysées<sup>88</sup> prennent en considération les deux compétences ou si elles n'en privilégient qu'une. A ces deux composantes, les épreuves sommatives du cycle moyen –et de tout le système éducatif- intègrent parfois une évaluation des ressources linguistiques (contenus grammaticaux) enseignées pour en mesurer la maîtrise par les apprenants. Cette évaluation des ressources linguistiques, nous la vérifierons à ce niveau.

- La nature des outils évaluatifs

Pour cet indicateur, nous déterminerons si les épreuves évaluent les compétences visées à travers des items, des situations ou si elles introduisent les deux outils en parallèle. Nous savons que les situations sont les outils les plus adaptés à une évaluation de compétences alors que les items sont privilégiés pour évaluer des apprentissages fragmentés. Déterminer la nature des outils investis pour évaluer chacune des composantes montre à quel degré les prescriptions de l'APC sont implantées dans les pratiques évaluatives.

- La manière dont les compétences sont évaluées

A travers cet indicateur, nous saurons si les deux compétences, dans le cas des épreuves évaluant les deux, sont évaluées séparément ou en parallèle à travers des situations complexes. Cet indicateur est inspiré de l'ouvrage de Huver et Springer (2011) qui rangent l'évaluation séparée des compétences dans les pratiques classiques qui ne correspondent plus aux prescriptions de l'APC. En effet, cette approche -qui rompt avec les pratiques anciennes- privilégie des situations qui évaluent indissociablement les deux compétences.

---

<sup>88</sup> Toutes les épreuves analysées sont écrites car l'évaluation demeure essentiellement écrite à tous les niveaux de l'enseignement scolaire. Certains enseignants pensent à tort que la note attribuée en évaluation continue représente les capacités des apprenants au niveau de l'oral, ce qui n'est pas le cas. En effet, la note de l'évaluation continue est attribuée en fonction de critères qui ne sont pas toujours liés à la performance orale comme la présence, la discipline, les devoirs réalisés...

Ces indicateurs retenus au niveau de cette première partie permettent de recueillir des données sur les compétences évaluées à travers les épreuves écrites analysées ainsi que sur les outils évaluatifs introduits, ce qui aboutit à une description structurale de ces épreuves. Dans ce qui suit, nous exposerons les indicateurs relatifs aux supports que les sujets analysés comportent pour en justifier la sélection.

### **2.1.3. Les supports**

Les supports regroupent tous les documents (textes, illustrations, schémas...) que les épreuves sommatives comportent pour fournir aux apprenants des informations, utiles ou inutiles, à exploiter dans la réalisation des tâches prescrites. La sélection de ces supports obéit à plusieurs critères dans le cadre de la nouvelle approche (APC) où des documents authentiques, variés, adaptés au niveau des apprenants... sont préconisés. Ce sont ces critères-ci que les indicateurs suivants nous permettent de vérifier :

- La nature du support proposé

Les supports que les épreuves contiennent varient entre textes, textes illustrés, tableaux, schémas... A travers cet indicateur, nous préciserons à quelle catégorie les enseignants recourent le plus pour évaluer les compétences de leurs apprenants. Certaines critiques pourraient réfuter cet indicateur étant donné que pour évaluer les compétences écrites en langues, les supports textuels seraient le choix le plus pertinent. Néanmoins, d'autres supports –iconiques en guise d'exemple- s'avèrent très utiles en évaluation de langues. A travers notre grille, nous ne comptons pas comparer les supports en vue d'en déterminer le plus utile : notre objectif se réduit à recenser les supports les plus investis en évaluation du FLE au cycle moyen. Pour représenter les résultats liés à cet indicateur, nous retenons les variables suivantes : texte, texte illustré et illustration, qui sont les trois supports les plus fréquents en langues.

- Le nombre de supports

Après avoir déterminé la nature des supports que les épreuves analysées comportent, nous en recenserons le nombre. Cet indicateur est utile étant donné que les situations évaluant les compétences doivent contenir un nombre de supports qui équivaut à celui des supports proposés dans les situations d'apprentissage et d'intégration. Trois variables sont retenues au niveau de cet indicateur : 1 support, 2 supports ou 3 supports.

- Les parties intégrant ces supports

Si les compétences sont évaluées séparément suivant le modèle classique inadapté aux exigences de la nouvelle approche actuellement appliquée, il devient nécessaire de préciser si

les supports sont sélectionnés dans la partie qui évalue la compréhension, celle qui évalue les ressources ou dans celle qui évalue la production. Cet indicateur est important car il fournira des éclaircissements sur l'exploitation des supports en évaluation du FLE et précisera si les situations évaluatives –quand elles sont introduites- contiennent des supports qui sont indispensables à leur validité.

- Le lien entre les thèmes des parties constituant les épreuves

Dans le cas des épreuves évaluant séparément les deux compétences, nous nous pencherons sur les thèmes retenus dans chaque partie. Notre objectif est de repérer d'éventuels liens rendant les supports sélectionnés en évaluation de la compréhension utiles à celle de la production. Les résultats obtenus à ce niveau sont répartis sur trois réponses éventuelles : *thèmes identiques*, *thèmes relevant du même domaine* et *thèmes différents*. Les thèmes sont considérés identiques si la consigne de la production propose un thème tout-à-fait similaire à celui de la compréhension. Si les deux thèmes sont complètement distincts, nous les considérons différents. Cependant, si les thèmes retenus –en dépit de leur différence- relèvent de la même thématique, ils seront considérés du même domaine.

- Des supports inédits

A travers cet indicateur, nous vérifierons si les supports proposés sont inédits, c'est-à-dire distincts de ceux travaillés dans les situations d'apprentissage. A ce niveau, notre intention est de nous assurer de la nouveauté des situations car si les supports textuels et iconiques sont déjà exploités, les situations évaluatives se transforment en activités de restitution, non conformes aux pratiques préconisées par la nouvelle approche. Mentionnons que les supports inédits sont ceux qui figurent dans les manuels scolaires des différents niveaux.

- Authentique, adapté ou fabriqué

Une autre précision est à apporter sur les supports sélectionnés. Il s'agit de déterminer si ceux-ci sont authentiques, adaptés ou fabriqués. Pour définir les premiers, nous nous référons à Cuq et Gruca qui précisent : « *les documents authentiques sont des documents « bruts », élaborés par des francophones pour des francophones à des fins de communication. Ce sont donc des énoncés produits dans des situations réelles de communication et non en vue de l'apprentissage* » (2005, p. 431). Cette définition des documents authentiques contient deux éléments qui les caractérisent. Le premier se rapporte au contexte de leur genèse qui est

authentique<sup>89</sup> et réel : ce ne sont pas des énoncés produits dans des situations simulées en vue de répondre à des objectifs précis. Ce sont ces objectifs qui fondent la deuxième caractéristique des documents authentiques qui ne sont pas produits à des fins pédagogiques. C'est grâce à leur caractère social que ceux-ci sont introduits pour présenter la langue étrangère apprise en situation. Contrairement à ces premiers documents originaux, les seconds sont des supports adaptés qui ont subi des changements en vue de répondre à des objectifs particuliers. Pour définir le texte adapté, nous citons Bentaifour qui avance :

C'est un texte qui a été retouché, c'est-à-dire qu'il a subi des transformations. Ces modifications portent sur le vocabulaire (des mots sont remplacés par d'autres), sur la syntaxe (des phrases sont reconstruites) et même sur la structure du texte (des passages entiers sont supprimés ou ajoutés) et elles sont justifiées par le désir de « rendre le texte plus abordable » (2009, p. 52).

Cette citation regroupe toutes les modifications que les enseignants apportent aux textes pour les rendre adaptés aux objectifs ciblés. Souvent, nous les reconnaissons à travers des mentions comme texte adapté, d'après...

La troisième catégorie de textes retenue regroupe les textes fabriqués, ou *didactiques* (Bentaifour, 2009). Selon cet auteur, ce sont des textes construits par les enseignants ou les auteurs des manuels en vue de cibler des apprentissages précis.

Dans notre étude, nous désignons par textes authentiques les textes dont les auteurs sont clairement cités ainsi que la source, les textes adaptés renvoient aux textes remaniés où les mentions « *adapté, d'après* » existent souvent, tandis que les textes fabriqués sont ceux rédigés par les enseignants. De ce fait, aucune référence ni mention - indiquant que le texte est adapté - ne figure (la mention « *texte élaboré* » pourrait cependant exister).

Les résultats de cet indicateur sont extrêmement significatifs car les approches récentes privilégient des *supports authentiques* représentant la langue dans son contexte réel pour un apprentissage plus efficace. Quant aux supports adaptés et fabriqués, leur introduction est déconseillée car ils pourraient modifier les idées et les expressions des auteurs, ce qui induit les apprenants en erreur.

- La source du texte

---

<sup>89</sup> La notion de document authentique ne suscite pas un consensus selon Cuq et Gruca (2005) étant donné que tout document du moment qu'il est investi en classe n'est plus authentique.

Cet indicateur découle du précédent et le justifie : la source permettra en effet de distinguer les textes authentiques des textes adaptés ou fabriqués. Trois variables représenteront les résultats liés à cet élément : source omise, source incomplète et source complète. Nous devons reconnaître que les résultats auxquels aboutira cet indicateur ne seront pas valides étant donné que certains enseignants ne citent pas fidèlement la source de leurs supports pour se conformer aux exigences : textes adaptés présentés comme authentiques, textes didactiques attribués à des auteurs connus...

- La longueur du texte

Nous ne disposons pas d'une norme définissant la longueur idéale du texte que les épreuves contiennent. C'est la raison pour laquelle nous nous référons aux programmes du cycle et retenons entre 10 et 15 lignes<sup>90</sup> comme une longueur acceptable des supports textuels sélectionnés<sup>91</sup>. Si nous insistons sur les textes à ce niveau, c'est parce que c'est l'unique catégorie où le problème de longueur se pose. Aussi, nous supposons que les enseignants de langue évaluent souvent les compétences écrites, notamment la compréhension, à travers des supports textuels, une hypothèse qui sera validée ou rejetée, grâce aux résultats du premier indicateur de cette partie de notre grille.

Contrairement aux indicateurs précédents, les résultats relatifs à celui-ci seront présentés par niveau vu les différences significatives constatées entre les textes sélectionnés dans les niveaux du cycle (1AM, 2AM, 3AM et 4AM).

En somme, tous les indicateurs retenus et exposés ci-dessus correspondent aux prescriptions de l'évaluation dans le cadre de l'APC. Il ne s'agit pas néanmoins d'une liste intégrale étant donné que les indicateurs retenus sont les plus adaptés à l'évaluation de la discipline dont il est question dans notre recherche, le FLE. Dans les points suivants, nous présenterons les indicateurs relatifs aux outils de mesure qui sont introduits par les enseignants du cycle.

#### **2.1.4. Les outils de mesure**

Pour nous conformer à la définition de validité antérieurement formulée, nous consacrons les deux parties restantes de notre grille aux outils de mesure et aux outils de

---

<sup>90</sup>Dans le « *Programme de français 3AM* », les concepteurs estiment que les supports sélectionnés par les enseignants doivent être d'une longueur d'environ 150 mots (de 10 à 15 lignes).

<sup>91</sup> Mentionnons que certains auteurs comme Bentaïfour (2009) préconisent la sélection de textes courts en classe de langue pour exploiter pleinement le texte. Cette recommandation, en dépit de son utilité incontestable, ne s'applique pas sur le cas de l'évaluation où les compétences évaluées sont antérieurement développées par les enseignants.

jugement en vue d'en vérifier la congruence avec les exigences de l'APC. Quels indicateurs notre grille retient-elle pour évaluer les outils de mesure analysés ? Nous devons mentionner que les outils évaluant les diverses composantes sont distincts, d'où la diversité des indicateurs retenus que nous exposerons sans respecter leur ordre dans la grille conçue. Constatons que certains indicateurs sont communs aux outils évaluant la compréhension et les ressources, nous les présenterons une seule fois dans ce qui suit.

- Le nombre de questions

Le nombre de questions proposées pour évaluer la compréhension et les ressources est à préciser, ce qui permettra de déterminer si des nombres fréquents de questions existent et si des instructions officielles les recommandent. De prime abord, ce critère paraît inutile à notre étude étant donné que les items constituent des outils évaluatifs inadaptés aux exigences de l'APC actuellement appliquée dans le système éducatif national ; en en déterminant le nombre, nous ne fournissons pas en effet des précisions sur la validité des épreuves analysées. Nous précisons, cependant, les nombres des questions évaluant la compréhension et les ressources pour les comparer et déterminer à quelle compétence l'évaluation du FLE au cycle moyen s'intéresse le plus.

- Les types des questions posées

Plusieurs items sont proposés pour une évaluation écrite des compétences des apprenants. Nous en déterminons le type pour recenser les types les plus récurrents dans les épreuves analysées. Certaines questions de chaque type seront exposées en vue de préciser si les enseignants tiennent compte des règles qui régissent la conception des divers types introduits. Notre grille énumère plusieurs types de questions évaluant la compréhension :

Les QCM : Dans ce type de questions, les apprenants sélectionnent une ou des réponses correctes parmi plusieurs solutions proposées. La classification établie par Leclercq (1986) les rangent -avec les *questions vrai/faux*- dans les sous-types des questions fermées. Quant à celle de Cuq et Gruca (2005), elle les présente comme un type autonome car seules les questions auxquelles nous répondrons par oui/non, vrai/faux sont intégrées dans les questions fermées. C'est à cette deuxième classification que nous nous référons dans notre étude où les QCM se présentent comme un type autonome.

Les questions Vrai/faux : Ce sont des questions où des phrases correctes et incorrectes sont proposées ; la tâche des apprenants consiste à distinguer celles qui sont correctes de celles qui ne le sont pas. Une variante de cette question consiste à exiger des apprenants la correction des phrases jugées incorrectes pour mieux vérifier leur compréhension. Contrairement à la



classification de Cuq (2005) qui les range dans les questions fermées, nous les présentons dans notre étude comme un type autonome.

Les questions ouvertes : Ces questions requièrent des apprenants une formulation des réponses ; et une mobilisation de leurs connaissances linguistiques.

Les questions orientées : Dans ce type, nous ne demandons pas aux apprenants de repérer des informations du texte, ni d'en vérifier la correction. Les informations leur sont exposées afin qu'ils les justifient et les éclaircissent (en s'appuyant sur les supports fournis).

Les questions fermées : Ces questions requièrent une réponse affirmative ou négative et nécessitent une certaine compétence linguistique pour une formulation correcte.

Les tableaux à compléter : Il est demandé aux apprenants de remplir les cases par des informations extraites du texte, tout comme les activités du *repérage d'informations*. Le nombre des informations à repérer constitue cependant une barrière qui sépare les premières des secondes où une seule information précise est à identifier dans les textes lus.

Les titres à donner : A travers ces questions, les apprenants doivent proposer un titre au texte lu. Mentionnons que les questions où des solutions sont proposées sont intégrées dans la catégorie des QCM.

Les activités de classement : La variante la plus courante consiste à demander aux apprenants de classer des éléments listés dans un tableau.

Ces activités, qui ne sont pas exhaustives, correspondent à celles que les enseignants introduisent souvent pour évaluer la compréhension. Pour évaluer les ressources, certaines activités sont similaires à celles précédemment citées (QCM, questions ouvertes...) alors que d'autres sont particulières à l'évaluation des ressources. Voici quelques-unes :

Identification : Les enseignants proposent aux apprenants des phrases souvent décontextualisées et leur demandent d'identifier des mots de nature ou de fonction précise. Proche de cette activité, l'activité de repérage est d'une complexité plus élevée étant donné que les mots à identifier sont à repérer du texte lu ; ce qui mobilise une attention plus importante ;

Réécriture : Dans cette catégorie, nous intégrons les questions qui demandent aux apprenants de transformer des phrases d'un nombre à un autre, d'un genre à un autre, d'un style à un autre, d'une forme à une autre... ;

Décomposition : Elle concerne les mots (identifier le préfixe, le radical et le suffixe) ainsi que les phrases (identifier les propositions, leurs natures et leurs fonctions)...

Les types cités ci-dessus ne sont pas exhaustifs car ils ne représentent que les activités récurrentes : certaines activités rarement repérables dans le corpus analysé sont regroupées sous la variable (*autres*) que contiennent les graphiques représentant les résultats obtenus.

Mentionnons que notre étude se limite à déterminer la validité des outils investis pour évaluer les compétences retenues, c'est pour cette raison que la correction de la formulation des consignes ainsi que leur précision ne seront pas abordées. Néanmoins, cette validité ne dépend pas seulement du nombre et des types des questions sélectionnées : les niveaux cognitifs visés par les évaluateurs sont à prendre en considération étant donné que l'évaluation des compétences, contrairement à celle des connaissances, doit mobiliser les niveaux supérieurs. Quels niveaux notre étude retient-elle ?

- Les niveaux cognitifs

Que les items évaluatifs proposés relèvent des niveaux cognitifs supérieurs est un des indicateurs de la validité des épreuves selon De Landsheere (2000). De ce fait, nous tentons de préciser le taux des épreuves évaluant les opérations cognitives complexes qui révèlent une maîtrise réelle des enseignements transmis. Pour classer les consignes par niveaux cognitifs, nous nous référons à Morissette (1993) et à Prigent (1990) qui listent les verbes et les contenus liés chacun des niveaux. Ainsi, les verbes *connaître, répéter, citer, réciter se rappeler, reconnaître, identifier, lister, nommer, énumérer, classer...* se rapportent à la **connaissance**<sup>92</sup>, premier niveau taxonomique. Le second niveau, la **compréhension**<sup>93</sup>, se traduit par des verbes comme *comprendre, saisir le sens, transposer, interpréter, reformuler, expliquer, illustrer, représenter...* Le niveau suivant (l'**application**<sup>94</sup>) correspond aux verbes : *utiliser, employer, appliquer, résoudre...* L'**analyse**<sup>95</sup>, quatrième niveau taxonomique, correspond aux verbes : *analyser, décomposer, expliquer, décrire, déduire, identifier* (les causes, les conséquences, les effets...), *déterminer, distinguer, inférer...* Pour le niveau de la **synthèse**<sup>96</sup>, des verbes comme : *formuler, rédiger, concevoir, élaborer, simuler, synthétiser, composer, planifier...* peuvent s'y

---

<sup>92</sup>**La connaissance**, le premier niveau taxonomique, consiste en la restitution de ce qui est mémorisé par les apprenants (faits, dates, règles de langue, informations, méthodes, théories...) (Morissette, 1993 ; Prigent, 1990).

<sup>93</sup>La compréhension consiste à expliquer ou expliciter les connaissances par des reformulations, des illustrations ; interpréter des expressions ; réduire un contenu aux informations essentielles ... (Prigent, 1990 ; Morissette, 1993).

<sup>94</sup>**L'application** correspond au transfert des connaissances théoriques dans des contextes concrets. A ce niveau, les apprenants doivent sélectionner des connaissances et des stratégies pour résoudre des problèmes (Prigent, 1990 ; Morissette, 1993). La connaissance, la compréhension et l'application sont des niveaux qui impliquent des activités intellectuelles élémentaires, contrairement aux niveaux suivants.

<sup>95</sup>**L'analyse** consiste à décomposer un tout à travers un raisonnement hypothético-déductif. Elle consiste alors à identifier les constituants ainsi que les liens qui les unissent (Prigent, 1990).

<sup>96</sup>**La synthèse** consiste à réunir et à agencer des éléments en vue de constituer un tout. Selon Prigent (1990), cette opération intellectuelle supérieure ne pourrait pas être réduite au résumé car elle requiert en effet la création d'éléments nouveaux.

rapporter. Enfin, le dernier niveau taxonomique, l'**évaluation**<sup>97</sup>, est représenté par les verbes suivants : *comparer, estimer, juger, évaluer, décrire, montrer, justifier, expliquer, distinguer...*

Pour certains verbes qui apparaissent dans des niveaux distincts, c'est le contenu de la consigne qui détermine avec précision le niveau cognitif auquel appartient chaque item.

Cet indicateur concerne les items qui évaluent les ressources. Pour ceux qui évaluent la compréhension, les niveaux suivants sont retenus :

- Niveaux de compréhension

Des modèles qui listent les niveaux de compréhension en lecture, deux semblent utiles à notre étude. Le premier est celui de Kintsch (1988) qui distingue trois niveaux de représentations sémantiques : le *niveau de surface*, la *base de texte propositionnelle* et le *modèle en situation*. Les questions du premier niveau interpellent des informations repérables littéralement dans le texte proposé, celles du deuxième niveau mobilisent les connaissances lexicales et syntaxico-sémantiques pour formuler des réponses, alors que celles du dernier niveau requiert une interprétation fondée sur les données du texte (Magherbi, Rocher, Gyselinck, Trosseille & Tardieu, 2009). Le second, conçu par Giasson (2000), liste quatre niveaux dont les trois premiers sont étroitement liés à ceux de Kintsch (1988). Cette auteure liste : la *compréhension littérale*, la *compréhension inférentielle* ou *interprétative*, la *compréhension critique* qui correspondent aux niveaux du premier modèle précédemment cité. A ces niveaux, s'ajoute la *compréhension créative* qui consiste à appliquer les représentations construites à sa vie personnelle.

Ce sont les niveaux de ce second modèle, plus détaillé que le premier, qui serviront de variables lors de la présentation des résultats liés à cet indicateur. Soulignons que la compréhension littérale, la plus superficielle, ne révèle pas une compréhension réelle des significations des supports lus.

Le dernier indicateur relatif aux questions évaluant la compréhension et les ressources est l'interdépendance des questions qui est déconseillée en évaluation pour des raisons dûment justifiées.

- Interdépendance des questions

Les items sélectionnés en évaluation sont interdépendants si la réponse à un item donné est conditionnée par celle de l'item précédent. De cette manière, nous aboutirons à évaluer

---

<sup>97</sup>L'*évaluation* consiste à juger, critiquer, évaluer des contenus, des méthodes, des théories... (Morissette, 1993 ; Prigent, 1990).

maintes fois un même élément, ce qui diminue considérablement la représentativité du contenu enseigné et sanctionne injustement les apprenants qui ne maîtrisent pas cet élément-ci. Aux questions interdépendantes, nous opposons les questions dépendantes qui respectent les exigences étant donné que les dérives citées sont écartées.

Tous ces indicateurs que nous venons d'exposer sont retenus pour étudier la validité des instruments mesurant les compétences compréhensive et grammaticale. Comme ces deux compétences ne sont pas les seules évaluées dans les épreuves de FLE en Algérie, notre grille investie inclut des indicateurs qui détermineront si les situations, étant les instruments de mesure récurrents en évaluation de la compétence expressive, répondent à toutes les exigences de validité. Ces indicateurs dont est constituée notre grille sont inspirés des travaux de De Ketele (1996) et Gérard (2009) sur les situations d'évaluation conformes aux exigences de l'APC.

- Types d'activités évaluant la production

Nous recensons à travers ce premier indicateur les types d'activités auxquelles les enseignants recourent le plus souvent pour évaluer la compétence expressive de leurs apprenants. Les résultats seront répartis sur les variables suivants, qui correspondent aux types récurrents :

La rédaction : la consigne proposée dans ce type demande aux apprenants de rédiger des textes pour expliquer, argumenter, informer ... Les énoncés produits par les apprenants seront corrigés par les enseignants pour évaluer leur compétence expressive.

La remise en ordre : des fragments (mots, phrases ou paragraphes...) en désordre sont proposés et la tâche des apprenants consiste à rétablir leur ordre pour obtenir un texte cohérent ; d'où leur appellation : puzzles. Pour ce faire, ils s'appuient sur des repères comme les signes de ponctuation, les articulateurs logiques et chronologiques, les substituts... ; ce qui complique le découpage du texte qui ne doit pas être ambigu.

Le résumé : dans ce type, les enseignants demandent aux apprenants de résumer le support textuel proposé de façon à en retenir les idées essentielles, bien que cette technique rédactionnelle ne soit pas développée chez les collégiens algériens.

Les textes à compléter : souvent, il est demandé aux apprenants de compléter une partie manquante du texte (*chapeau, introduction, développement, conclusion...*) ; ce qui correspond au « texte à gouffre », expression introduite par Bouchard et reprise par Cuq et Gruca (2005).

Les textes à trous : ce type consiste à proposer aux apprenants un texte à trous, c'est-à-dire un texte dans lequel des mots –du même champ lexical ou de la même catégorie

grammaticale- sont supprimés. La consigne qui leur est donnée est de remplir ces trous par les mots appropriés, ce qui requiert une bonne compréhension du texte retenu. La nature des éléments supprimés (mots, phrases, signes de ponctuation...) et leur nombre dépend des objectifs des examinateurs. Ce type est souvent confondu avec *le test de closure* où la suppression des mots est aléatoire étant donné que ceux-ci n'appartiennent pas nécessairement à la même catégorie grammaticale. Selon Gagatsis (1985), qui cite W. L. Tylor (1953)-, Reinwein (1987) et De Landsheere (1992), les examinateurs procèdent à la suppression régulière des 5ème mots (1er, 6ème, 11ème... ; 2ème, 7ème, 12ème...) que les apprenants doivent restituer. Soulignons que ces intervalles pourraient être diminués pour la complexité de la tâche.

Ainsi, quelques différences entre les deux types pourraient être repérées. La première est liée à l'intervalle des mots supprimés : régulier dans le test à closure, il ne l'est pas dans le texte à trous. La deuxième se rapporte aux mots supprimés : contrairement au test de closure où les mots sont omis aléatoirement, le texte à trous efface des mots particuliers dans le texte (connecteurs, verbes...). A ces deux différences, s'ajoute la compétence visée : alors que le test à closure qui ne cible pas une compétence précise et intègre une variété de connaissances, le texte à trous évalue la compréhension à travers une compétence précise (grammaire, orthographe...).

Bien que *les textes à trous* tels que nos épreuves sommatives les introduisent ne correspondent pas réellement à leur description théorique citée ci-dessus et empruntée à Cuq et Gruca (2005), nous retenons cette appellation très courante dans le milieu scolaire.

**La réécriture** : dans ce type, les évaluateurs demandent aux apprenants de réécrire le support textuel sélectionné pour obtenir des dialogues, des BD...

**Aucun** : cette variable correspond aux épreuves qui ne proposent aucune activité pour évaluer la compétence expressive.

Ce sont les résultats que nous exposerons dans le chapitre suivant qui nous révéleront les types les plus fréquents dans les épreuves sommatives constituant notre corpus. De tous ces types, seules les situations de rédaction qui aboutissent à des productions écrites concrètes seront soumises à une évaluation pour en déterminer la validité, et ce, à travers les indicateurs suivants qui sont liés à certaines caractéristiques des situations évaluatives idéales.

- Situations inédites ?

Dans une évaluation de compétences, les situations d'évaluation proposées doivent être inédites ou nouvelles, c'est-à-dire des situations qui ne sont pas déjà travaillées en classe.

Soulignons que ces situations inédites, dont les paramètres (nombre de supports, types des textes...) sont identiques à ceux des situations issues de la même famille, mobilisent les ressources antérieurement apprises qui seront articulées différemment. Parmi les situations évaluatives sélectionnées par les enseignants du cycle moyen pour évaluer leurs apprenants en FLE, lesquelles seront considérées inédites ?

Les situations inédites dans notre étude sont celles qui ne figurent pas dans les manuels scolaires des niveaux du cycle. Notre recours aux manuels scolaires pour séparer les situations nouvelles de celles qui ne le sont pas s'impose vu que les situations qui y sont proposées sont celles que les enseignants introduisent souvent dans leurs classes. En effet, le manuel scolaire est toujours un outil précieux aux yeux de nos enseignants qui continuent à exploiter ses supports, ses activités, ses situations...

Les résultats auxquels cet indicateur aboutit permettront de déterminer si les enseignants recourent à des situations inédites comme l'évaluation dans le cadre de l'APC le requiert, ou si les situations travaillées qui ne sont en réalité que des activités de restitution vivement déconseillées dans une logique de compétences. Que les situations évaluatives soient inédites ne constitue pas le critère unique auquel les enseignants doivent tenir compte de la conception et de la sélection de ces situations-ci. En effet, celles-ci doivent être complètes, ce qui nécessite le respect de plusieurs indicateurs que nous listerons ci-dessous.

- Clarté ou précision du contexte

Pour être complètes, les situations évaluatives doivent contenir certains constituants essentiels comme le contexte, les supports, les tâches... Le premier constituant, le contexte, doit éclaircir l'environnement de la tâche, la fonction des productions attendues... Ce sont ces paramètres qui permettent aux apprenants de produire des énoncés adaptés à la situation et aux enseignants d'en évaluer la cohérence.

Lors de la conception des situations, un énoncé précédant les consignes est à insérer pour fournir aux apprenants les précisions nécessaires sur les situations communicatives proposées. Dans notre étude, nous considérons clair le contexte lorsque la situation évaluative fournit au moins une des précisions nécessaires.

- Ce que le contexte précise...

Comme mentionné précédemment, le contexte pourrait préciser *l'environnement* de la tâche, la *fonction de la production* et le *destinataire*. Ces éléments constituent les variables retenues pour représenter les résultats liés à cet indicateur qui nous permettra de comprendre si les situations conçues par les enseignants du cycle privilégient certains éléments du contexte.

Mentionnons que par environnement, nous désignons le thème à traiter par les apprenants, le cadre spatio-temporel, les circonstances... ; éléments qui déterminent les idées à insérer dans les productions, niveau de langue... Quant au destinataire, il renvoie aux personnes à qui ces productions sont destinées ; ce qui contribue à adopter un niveau de langue adapté à la situation. Pour la fonction de la production, elle explicite pour quoi les énoncés sont produits ; ce qui en détermine la visée.

- Supports proposés

Cet indicateur et celui qui le suit sont liés au deuxième constituant essentiel des situations évaluatives : les supports. Nous préciserons en premier lieu si les situations conçues par les enseignants de FLE contiennent des supports qui fournissent aux apprenants quelques informations utiles à la résolution des situations proposées. Rappelons que les supports fournis ne se réduisent plus aux supports textuels dans la mesure où des illustrations, des tableaux, des schémas...pourraient être également sélectionnés.

- Utilité des supports sélectionnés

A travers cet indicateur, nous déterminerons si les supports sélectionnés sont utiles. De prime abord, les résultats liés à cet indicateur semblent subjectifs étant donné que cette utilité ne pourrait être aisément mesurée. Cependant, en associant celle-ci à la capacité des supports sélectionnés de fournir des renseignements utiles à la résolution des situations-problèmes, la tâche devient plus aisée. Il est important de préciser à ce niveau de notre étude que les supports sélectionnés doivent alterner les informations utiles et inutiles. En effet, les apprenants -pour investir ces informations dans leurs productions, procèdent à une sélection minutieuse pour ne repérer que ce qui est pertinent.

- Respect de la règle des 2/3

Cet indicateur correspond au nombre des tâches proposées dans les épreuves évaluatives du cycle que nous devons vérifier pour préciser si la règle des 2/3 est respectée. En effet, les situations ou épreuves analysées, pour être conformes aux exigences de l'évaluation des compétences, devraient contenir au moins trois tâches. Selon De Ketele (1996) et Gérard (2009), la compétence est déclarée maîtrisée si elle est réussie dans deux occasions sur trois ; c'est la règle des 2/3, abondamment expliquée dans notre volet théorique.

Dans notre étude, nous rangerons les situations qui proposent trois tâches dans la catégorie des épreuves qui respectent la règle évoquée ; celles qui ne contiennent pas le nombre requis seront classées dans la catégorie des épreuves non-conformes aux exigences.

- Interdépendance des tâches

Cette notion d'interdépendance –évoquée supra- s'applique également aux situations évaluatives dont les trois tâches prescrites doivent être indépendantes. Tout comme les items interdépendants sont déconseillés pour ne pas sanctionner plusieurs fois la maîtrise d'un élément précis, les tâches interdépendantes le sont. Si les situations proposées par les enseignants contiennent plus d'une tâche, il serait nécessaire de vérifier si chaque nouvelle tâche ne dépend pas de celle qui la précède. Cette interdépendance n'est pas en effet garante de la validité des évaluations étant donné que nous ne proposons pas, dans ce cas-ci, plusieurs tâches pour vérifier la maîtrise de la compétence ; mais plutôt une *tâche à tiroirs*<sup>98</sup>.

- Précision des instructions de réalisation

Les deux derniers indicateurs liés aux outils de mesure concernent les instructions de réalisation<sup>99</sup> qui renvoient, selon Gérard, à « *tous les éléments matériels* » (2009, p.78) de la situation. Nous préciserons en premier lieu si ces instructions sont introduites par les enseignants dans les situations que contiennent les épreuves analysées. Si la situation propose au moins une des précisions éventuelles, nous la rangeons parmi les situations qui contiennent des instructions de réalisation.

- Ces précisions concernent...

Pour une meilleure description des situations évaluatives conçues par les enseignants algériens du cycle moyen, nous préciserons si les instructions de réalisation se réduisent à celles listées par Gérard (2009), détaillées dans notre volet théorique, et si les enseignants incluent des éléments nouveaux. Les variables que nous retenons pour présenter les résultats sont :

Les acquis à mobiliser : ce sont les outils linguistiques (ressources) que les apprenants doivent investir pour résoudre les situations-problèmes sélectionnées. Ces outils linguistiques concernent : le vocabulaire à investir, les temps verbaux appropriés, les types et les formes de phrases...

---

<sup>98</sup> Nous proposons cette notion que nous formulons sur le modèle des *questions à tiroirs* pour signifier les tâches où la résolution de chacune est conditionnée par la résolution de celle qui la précède, tout comme les questions à tiroirs où chaque réponse est fondée sur la réponse qui la précède (Gérard, 2009).

<sup>99</sup> Gérard (2009) se sert du terme consigne pour désigner les instructions de travail qui englobent des précisions sur la nature de la production, sa longueur, sa présentation, la méthode de réalisation (individuelle ou collectif)...



La procédure de réalisation : c'est la démarche à emprunter pour la réalisation des tâches. Elle concerne le plan du texte à produire, les étapes de réalisation...

Les données techniques : elles se rapportent à la présentation du texte (disposition en paragraphes, en colonnes...), à sa longueur...

Cet indicateur et les indicateurs précédents retenus pour étudier la validité situations évaluative évaluant la compétence expressive des collégiens algériens en FLE se rapporte essentiellement à deux caractéristiques : la nouveauté et la complétude. Ce sont des caractéristiques que nous pourrions vérifier objectivement à travers des indicateurs précis, contrairement aux autres caractéristiques (complexité, authenticité...) dont les résultats ne seront pas exempts de subjectivité.

Ainsi, nous clôturons la troisième partie de notre grille qui porte sur les indicateurs liés aux instruments que les enseignants introduisent pour évaluer les compétences des apprenants en FLE. Comme la validité des épreuves évaluatives ne se limite pas à celle des instruments de mesure, nous étudierons également les instruments de jugement investis à travers plusieurs indicateurs que nous exposerons dans les éléments qui suivent.

#### **2.1.5. Les outils de jugement**

Les indicateurs présentés dans cette catégorie sont relatifs aux instruments de jugement investis pour évaluer les réponses des apprenants. Pour proposer des éléments adaptés à la particularité de ces outils, nous nous référons aux informations de notre volet théorique. Certes, nous détaillerons essentiellement deux instruments que nous jugeons plus adaptés à la perspective sommative en évaluation : le barème de notation et la grille critériée ; la liste de vérification étant plus utile en évaluation formative.

- Instrument investi

A travers cet indicateur, nous déterminerons les instruments investis dans les épreuves écrites : les barèmes de notation, les listes de vérification ou les grilles critériées ; trois outils qui pourraient être investis dans une perspective sommative. Cette précision est nécessaire car elle définit à quel(s) instrument(s) les enseignants recourent le plus et si celui-ci est adapté à une évaluation rigoureuse des compétences évaluées. Elle permet également de vérifier si notre évaluation sommative n'exclut pas la liste de vérification en tant qu'instrument de jugement.

Les résultats seront ainsi répartis sur les variables suivantes : le *barème de notation*, la *liste de vérification*, la *grille critériée* et *aucun* (les épreuves où aucun outil de jugement n'est investi).

Lorsque le barème de notation est l'instrument de jugement investi, les indicateurs suivants sont à étudier :

- Parties où le barème est investi

Lorsque les épreuves évaluatives évaluent séparément la compréhension, la production et les ressources linguistiques, il devient important de vérifier à quelle partie cet outil de jugement est appliqué. Cet indicateur est important car il nous permet de juger si cet outil est approprié aux outils de mesure investis dans les épreuves constituant notre corpus. Si les items sont les instruments de mesure investis par les enseignants, le barème est un outil de jugement approprié. Cependant, il ne l'est pas dans le cas des compétences évaluées à travers des situations.

- Taux accordé aux ressources

Quand les notes partielles sont mentionnées, nous déterminerons le taux accordé aux ressources (dans le cas des épreuves qui les évaluent séparément). Cet intérêt que nous lui accordons s'impose vu les travaux théoriques menés par des auteurs comme Gérard (2009) qui estiment que la note attribuée aux ressources linguistiques ne doit pas dépasser 25% de la note totale. Les résultats auxquels nous aboutissons révéleront à quel degré cette exigence est respectée dans les épreuves sommatives du cycle moyen. Ils seront répartis sur les variables suivants : *moins de 25%*, *de 26% à 50%* et *plus de 50%*.

Cet indicateur paraît -à première vue- contredire nos affirmations antérieures selon lesquelles une perspective sommative exclut une évaluation de ressources. Son introduction cependant est nécessaire car des enseignants continuent à les évaluer dans les épreuves conçues.

Ces deux indicateurs se rapportent à la validité des barèmes de notation. Dans le cas des épreuves exploitant les grilles de correction, des critères distincts seront investis.

- Parties où la grille critériée est investie

Si des grilles de correction sont investies, nous déterminerons à quelles parties elles sont appliquées (dans le cas d'une évaluation séparée des compétences), ce qui nous permettra de préciser leur compatibilité avec les outils de mesure investis. Rappelons que, dans une perspective de compétences, les grilles critériées constituent les outils de jugements les plus adaptés pour évaluer les productions des apprenants.

- Clarté des grilles

Par clarté des grilles de correction, nous entendons la présence de critères indiquant les points de vue adoptés pour la correction des productions, la présence d'indicateurs explicitant chaque critère et la précision de l'échelle de mesure.

- Précision des critères

La grille critériée contient un nombre de critères qui doivent être pertinents et précis. Leur précision est fondée sur la clarté des *éléments observables* ou *indicateurs* qui permettent de mieux identifier chacun des critères sélectionnés. Elle requiert, en effet, une formulation claire décrivant de manière détaillée chaque critère de façon à ce que tout enseignant, même s'il n'est pas le concepteur de l'épreuve, peut corriger les productions des apprenants sans difficultés.

- Précision des indicateurs

Les indicateurs, qui constituent la concrétisation des critères dans les productions corrigées, doivent être précis et pertinents. Leur pertinence comme celle des critères est garantie d'une correction rigoureuse et équitable.

- Indépendance des critères

Les critères doivent être indépendants : aucun critère ne doit être lié au précédent ni au suivant. Cette règle assure une évaluation de plusieurs aspects des productions des apprenants : si les critères sont liés ; les correcteurs corrigeront un seul aspect de ces productions négligeant d'autres aspects aussi importants.

- Clarté des critères minimaux et de perfectionnement

Une distinction est établie entre les critères minimaux et les critères de perfectionnement. Cette distinction, éclaircie dans le volet théorique, doit être explicite dans les grilles conçues par les enseignants.

- Pertinence des types de critères

Les types de critères investis doivent être pertinents, adaptés à la particularité de la discipline. En effet, chaque discipline nécessite des critères minimaux et de perfectionnement précis. Ainsi, le critère « *correction de la langue* », qui est un critère de perfectionnement dans plusieurs disciplines, ne peut pas l'être en FLE où la correction linguistique est une condition nécessaire pour la maîtrise de la compétence.

- Nombre des critères

Certes, les critères de correction retenus ne doivent pas être nombreux. Ceux-ci ne doivent pas dépasser quatre critères selon Gérard (2009) : trois critères minimaux et un critère de perfectionnement. Cette exigence est-elle respectée dans les grilles critériées adoptées ?

- Taux accordé aux critères de perfectionnement

Les critères de perfectionnement ne doivent pas dépasser les  $\frac{1}{4}$  de tous les critères retenus pour corriger les productions des apprenants (Gérard, 2009). Cet indicateur est à vérifier dans le cas où les grilles conçues par les enseignants explicitent les critères minimaux et de perfectionnement. En effet, dans le cas échéant, identifier la valeur de chaque critère devient une tâche difficile.

- Clarté des échelles de mesure

L'échelle de mesure constitue un élément essentiel de toute grille. Elle doit être claire et précise dans la grille pour des résultats valides reflétant les compétences des évalués. Rappelons que les épreuves écrites se soldent par des notes qui demeurent des pratiques privilégiées car « *elles se prêtent facilement à des opérations d'addition, de multiplication (coefficients) et de division (moyennes)* » (De Vecchi, 2011, p. 66), d'où la nécessité d'opter pour les échelles numériques.

En somme, les indicateurs que nous venons de citer nous semblent les plus indispensables pour assurer la validité de la correction et du jugement des productions des apprenants. Certes, ils ne relèvent que de deux instruments (le barème et la grille) qui sont les plus utilisés en évaluation sommative des apprentissages. En appliquant ces indicateurs et les autres dont notre grille est constituée, nous pouvons parvenir à quelques précisions sur les pratiques évaluatives des enseignants au cycle moyen en évaluation du FLE. C'est la description de ces pratiques qui permettra de repérer les failles de notre évaluation sommative pour proposer ultérieurement des pratiques plus conformes aux exigences de l'APC.

Nous devons mentionner à ce niveau de notre étude que nous exploiterons la même grille pour la comparaison des BEM qui précèdent l'APC à ceux qui la suivent, en opérant de légères modifications. En effet, les trois indicateurs qui suivent seront supprimés pour les raisons que nous exposerons :

- Des supports inédits : en vue de vérifier la nouveauté des supports sélectionnés par les évaluateurs dans les BEM évaluant les compétences des apprenants en FLE, nous devons recourir à des manuels scolaires très anciens qui étaient exploités durant les années des épreuves analysées (1997, 1998...). Comme nous n'avons pas pu les collecter, nous avons éliminé cet indicateur.

- Des situations inédites : tout comme l'indicateur précédent, la vérification de celui-ci requiert le recours aux manuels anciens que nous n'avons pas pu trouver.
- Types d'activités évaluant la production : dans les examens officiels, la rédaction est une activité indispensable pour évaluer la compétence expressive des apprenants.
- Constituants des situations d'évaluation : la complétude des situations d'évaluation évoquée de manière détaillée dans notre grille analytique sera étudiée à travers la vérification de leurs constituants essentiels : *contexte, supports, tâches, instructions*. Nous vérifierons alors si ces constituants essentiels existent dans chacune des situations proposées pour évaluer la production.

La grille dont nous venons de préciser les indicateurs ne constitue pas le seul instrument de collecte introduit dans notre recherche. Nous proposons une deuxième grille qui n'a pas le même objectif que la première. Elle vise la validation des dispositifs que nous proposerons dans cette deuxième partie de notre étude. Nous la présenterons succinctement dans les éléments qui suivent.

## 2.2. La liste de vérification

Cette deuxième grille se présente comme une liste de vérification conçue pour valider les situations évaluatives que nous concevons. Elle pourrait également assurer la validation des situations que les épreuves analysées comportent. Cependant, l'analyse des épreuves -qui constituent notre corpus- à travers des critères et indicateurs similaires à ceux de la liste de vérification est suffisante.

Nous devons mentionner que la conception de cette liste est inspirée de plusieurs travaux sur les situations d'évaluation valides. Ainsi, Gérard (2009) propose dans son ouvrage « *Evaluer les compétences-Guide pratique* » un outil de vérification pour valider les situations évaluatives : il s'agit d'une grille descriptive présentant toutes leurs caractéristiques. Quant à Roegiers (2006), il présente les qualités requises dans une épreuve d'évaluation. Ces qualités sont similaires à celles proposées par Gérard (2009), qui affirme que sa grille est adaptée de Roegiers (2003). La troisième référence qui a inspiré notre grille est le « *Document d'accompagnement du programme de français- 1<sup>ère</sup> AM* », paru en 2010, qui énumère des questions que les enseignants devraient se poser pour vérifier la validité des situations conçues.

Notre liste de vérification dont nous exposons les indicateurs s'inspire certainement de tous les travaux ci-dessus cités. Cependant, elle s'avère notamment comme une adaptation de la liste telle que Gérard (2009) la propose. Quels critères et indicateurs faut-il retenir alors ? Les critères -dont notre liste de vérification est constituée- sont en réalité les caractéristiques

qu'ont les situations évaluatives. Chaque caractéristique est explicitée en éléments concrets représentant les manifestations des critères dans les situations validées. Pour la vérification des indicateurs retenus, nous retenons deux alternatives (oui) ou (non). Dans le cas où certains indicateurs s'avèrent insuffisants, une réécriture de la situation s'impose pour qu'elle s'adapte aux exigences.

- La situation est-elle complexe ?

Nous considérons complexe toute situation qui présente un problème à résoudre, investit des ressources déjà enseignées et mobilise des opérations cognitives de haut niveau. Si un des indicateurs listés manque, la situation ne sera pas complexe dans la mesure où elle ne remplit pas toutes les conditions.

- La situation est-elle signifiante ?

Est considérée signifiante toute situation qui s'adresse directement aux apprenants pour les impliquer, constitue un défi à surmonter, porte sur un contexte familier qui relève de leur vie quotidienne et assure le transfert de ce qui est appris, rendant par-là significatifs tous les apprentissages réalisés tout au long de la séquence et du projet.

- La situation est-elle authentique ?

Nous jugeons authentique toute situation qui traite un sujet social, présente un contexte vraisemblable, propose des tâches réalistes exposant les apprenants aux contraintes de la vie réelle et aboutit à une production socialisable.

- La situation est-elle complète ?

Les composantes de la situation évaluative complète, nous les devons à Gérard (2009). Ce même auteur expose dans sa liste de vérification de nombreux indicateurs que nous reprenons. La situation complète est une situation qui mentionne clairement la production concrète et évaluable à laquelle aboutit la consigne. A ce niveau, la consigne doit préciser le but de la précision.

Nous vérifions également si toutes les composantes existent dans la situation conçue : le contexte (qui doit être clair), les supports (dont le thème, le domaine et le nombre dépend de certains critères), les tâches (dont la visée ainsi que le nombre sont à prendre en considération) ainsi que les critères de correction (qui doivent être sélectionnés en tenant compte des exigences).

- La situation est-elle inédite ?

Pour être inédite, la situation doit proposer des tâches nouvelles et des supports inédits. Pour cela, il suffit de s'assurer que ce qui est proposé n'est pas déjà abordé en classe.

- La situation est-elle adéquate ?

Contrairement aux éléments précédents puisés dans la grille de Gérard (2009), ce dernier est ajouté des qualités énumérées par Roegiers (2006). A travers celui-ci, nous vérifions si la situation évalue les compétences visées dans le programme tout en respectant les valeurs prônées par l'enseignement/apprentissage du FLE et si elle est adaptée au niveau des évalués. A ces deux indicateurs, s'ajoute le vocabulaire investi pour sa formulation qui doit éviter toute ambiguïté sémantique ou lexicale.

Tous ces éléments que nous reprenons nous servent à valider les situations que nous proposons. Ils seront également appliqués par les experts (enseignants et inspecteurs) que nous solliciterons pour recueillir des jugements sur ces situations-là. Toutes les failles qui jaillissent suite à l'application de cet outil de vérification indiquent que les situations sont à retravailler. Quant aux grilles de correction, nous adopterons les mêmes indicateurs de notre grille analytique pour les valider.

### 3. Contexte de l'étude

Notre recherche porte sur la validité de l'évaluation du FLE au cycle moyen, qui est un cycle transitoire entre le primaire et le secondaire. Dans les trois cycles, la langue française –à l'instar de la deuxième langue étrangère enseignée- est enseignée comme un outil de communication, de documentation et d'accès à des cultures différentes. En vue d'atteindre ces finalités, le système éducatif algérien a adopté des pédagogies et des approches variées. Depuis 2003, l'APC fait son entrée dans la scène éducative nationale et bouleverse les pratiques enseignantes et pédagogiques. Selon Adel Farid (2005), l'APC a abouti au changement des rôles des acteurs impliqués (enseignants et apprenants), à l'élaboration de nouveaux programmes scolaires -qui requièrent de nouvelles attitudes et compétences enseignantes- et à la précision des modalités d'enseignement et d'évaluation.

Tous ces changements opérés par la nouvelle approche visent l'amélioration de la qualité de l'enseignement (El Mistari, 2013). Ils sont le résultat des « *nombreuses critiques adressées à l'éducation sur l'absence de qualité* » (Hassani, 2013). Dans ce contexte, nous mentionnons que la PPO est considérée comme la source effective de l'échec constaté, un échec dû essentiellement à une présentation fragmentée des apprentissages sélectionnés (Ait Amar Meziane, 2014), qui ne favorise par leur intégration dans les situations réelles.

La mise en œuvre de l'APC pour résoudre certains problèmes scolaires constatés, comme l'inadaptation des contenus et les taux faibles de réussite (Tawil, 2005), n'était pas chaleureusement accueillie par les enseignants habitués à leurs anciennes pratiques. Dans ce contexte, Hassani (2013) évoque un écart entre les instructions dictées par la nouvelle approche

et les pratiques réelles des enseignants, un écart dû notamment à la réticence des enseignants. Ceux-ci ont dû s'adapter à une nouvelle ère d'enseignement à travers des situations complexes présentant la langue dans un contexte vraisemblable. Ces situations dites d'apprentissage permettent aux apprenants la découverte des ressources à mobiliser pour la production d'énoncés relevant du même type. Après la maîtrise des ressources, des situations d'intégration sont proposées pour que les apprenants s'entraînent à mobiliser ce qui est appris. Au terme des apprentissages, des situations d'évaluation donnent un aperçu sur la capacité des apprenants à transférer ce qui est appris dans de nouveaux contextes.

Ce rappel des pratiques liées à l'APC attribue un statut à l'évaluation qui est effectuée également à travers des situations complexes. Étant donné que l'évaluation constitue le thème de notre recherche, nous tentons de fournir quelques précisions sur l'évaluation dans le cadre de l'APC telle que les instructions officielles l'abordent. Pour ce faire, nous nous référons aux programmes, guides et documents d'accompagnement du cycle moyen<sup>100</sup> ; ainsi qu'à certains travaux sur la réforme de l'éducation nationale.

### 3.1. A propos des types d'évaluation

Tous les programmes, guides et documents d'accompagnement consacrent quelques éléments à l'évaluation du FLE en vue d'en expliciter les modalités et les pratiques. Les formes évaluatives introduites sont également abordées. Le « *Programme de français 3AM* » (2012-2013) évoque les trois formes les plus connues : évaluation diagnostique, évaluation formative et évaluation certificative, un terme préféré à celui d'évaluation sommative qui ne figure pas dans ce document. Le « *Programme de français de 1<sup>ère</sup> année moyenne* » (2010) évoque les mêmes formes en les situant aux différents moments du projet pédagogique : avant le projet, tout au long de celui-ci et à sa fin. Notons que d'autres formes d'évaluation comme l'auto-évaluation, existantes dans les manuels scolaires, sont évoquées, mais brièvement. Ainsi, le Document d'accompagnement au programme de 3AM (2012-2013) aborde succinctement l'auto-évaluation ainsi que certains outils qui aident à l'assurer.

Les trois formes évoquées ci-dessus sont introduites par les enseignants sans leur accorder le même statut. L'évaluation diagnostique est considérée peu utile par certains car les enseignants sont contraints à appliquer des programmes et des contenus préétablis, qui ne correspondent pas toujours aux besoins effectifs de leurs apprenants. Elle devient ainsi une des

---

<sup>100</sup> Nous devons mentionner que nous ne menons pas une étude synchronique pour retracer l'évolution des exigences de l'APC en matière de l'évaluation du FLE. Nous nous limitons à citer les exigences évoquées dans quelques documents recueillis.



tâches que certains enseignants accomplissent à contrecœur en recourant à des activités variées. En effet, certains enseignants recourent à des exercices grammaticaux pour dresser le profil d'entrée de leurs apprenants alors que d'autres optent pour des évaluations identiques aux évaluations sommatives pour un aperçu plus précis sur les acquis du public concerné.

Nous devons préciser que les enseignants ne disposent pas de renseignements suffisants sur cette évaluation et sur ses modalités. A ce niveau, certaines pratiques convergentes (nature des activités, communication des résultats...) illustrent cette situation, une convergence qui est dû essentiellement au nombre très restreint des travaux qui s'intéressent aux représentations et aux pratiques à adopter en évaluation diagnostique.

Quant à l'évaluation formative, elle est intrinsèque aux interventions enseignantes. Les enseignants l'intègrent inconsciemment, à travers les corrections simultanées aux erreurs commises par les apprenants, et consciemment à travers des activités variées révélant les lacunes auxquelles il faut remédier. Comparée à la première forme, cette évaluation est privilégiée dans l'univers scolaire car les instructions des inspecteurs préconisent une intervention régulatrice de toutes les incorrections constatées et une vérification régulière de la maîtrise des apprentissages. Pour ce faire, les enseignants prévoient des exercices et des activités tout au long des séquences pédagogiques<sup>101</sup> pour repérer les lacunes qui nécessitent une remédiation. Pour les leçons portant sur les ressources retenues, des exercices variés évaluant les savoir et les savoir-faire sont proposés pour l'automatisation de ce qui est enseigné.

A ces exercices consolidant l'apprentissage des ressources, s'ajoutent des activités à travers lesquelles les apprenants s'entraînent à intégrer ce qui est appris en contexte. Ce sont les activités de production, de préparation à la production et du compte-rendu des énoncés rédigés, qui permettent un apprentissage guidé<sup>102</sup> de l'intégration des ressources pour la résolution de situations-problèmes. Cependant, nous ne pouvons affirmer que les enseignants introduisent rigoureusement cet apprentissage de l'intégration dans la mesure où nous ne disposons pas d'études révélant les pratiques concrètes des enseignants en classe de FLE. En effet, les références consultées qui s'appuient sur des études théoriques décrivent et prescrivent ce qui doit être fait. Ces descriptions et prescriptions seraient néanmoins plus utiles si elles

---

<sup>101</sup> Les apprentissages dans le cadre de la pédagogie du projet sont structurés par projets aboutissant à des réalisations concrètes. Au cycle moyen, les apprentissages de chaque année scolaire sont organisés en trois projets constitués chacun de séquences (de deux à trois). La **séquence** est constituée de plusieurs activités (compréhension orale, expression orale, compréhension de l'écrit, ressources et production écrite).

<sup>102</sup> Par cet **apprentissage guidé**, nous entendons la présence du guidage de l'enseignant tout au long du processus étant donné que les apprenants ne pourraient investir les ressources que de manière progressive avant d'atteindre une certaine autonomie.

étaient fondées sur une approche praxéologique analysant les pratiques existantes pour la conception de nouveaux savoirs.

L'évaluation sommative est la dernière forme indispensable que les enseignants sont obligés d'introduire. En effet, les résultats auxquels elle aboutit ont un rôle incontournable dans les décisions prises (réussite ou échec, récompense ou sanction ...). Quels sont les évaluations qui relèvent de cette forme sommative ?

Au cycle moyen, à l'instar des autres cycles, plusieurs épreuves sont introduites tout au long de l'année scolaire qui est répartie en trois trimestres. Chaque trimestre contient trois évaluations sommatives : deux devoirs et une composition. Le devoir consiste en une épreuve écrite introduite au bout d'une séquence d'apprentissages pour en évaluer la maîtrise. Quant à la composition, elle clôt les apprentissages cumulés tout au long d'un projet pédagogique pour les évaluer. Dans les devoirs et les compositions, les compétences de compréhension et de production sont jugées à travers des activités variées que nous ne détaillons pas à ce niveau de notre recherche. Considérer les compositions comme des évaluations sommatives pourrait être aisément admis, contrairement aux devoirs qui correspondent aux fins des séquences. Néanmoins, si nous tenons compte des affirmations de Cuq (2003), nous réalisons que les évaluations sommatives correspondent à toutes les épreuves qui dressent le bilan des acquis, comme c'est le cas des devoirs.

Étant donné que notre recherche étudie essentiellement les épreuves sommatives du FLE au cycle moyen, une précision doit être ajoutée à ce niveau : l'absence des épreuves évaluant la compétence orale des apprenants. En effet, l'évaluation demeure écrite à tous les niveaux du cycle, ce qui pourrait révéler une certaine priorité accordée au développement des compétences écrites. Cette priorité doit être dépassée grâce à l'application de la nouvelle approche (APC) qui accorde le statut mérité à chacun de ces domaines complémentaires. Ceci ne correspond pas cependant à la réalité du milieu scolaire où l'évaluation de la compétence orale n'existe plus même si elle est aussi développée que la compétence écrite. Certains enseignants croient à tort que la note de l'évaluation continue<sup>103</sup> renvoie à celle de la compétence orale, ce qui n'est pas fondé étant donné que cette note est attribuée sur des travaux qui ne révèlent pas toujours la performance orale des apprenants.

---

<sup>103</sup> Dans les bulletins des apprenants, une case est réservée à la note de l'évaluation continue qui est aussi importante que celles des devoirs et compositions. Cette note est la somme de notes partielles attribuées sur plusieurs critères tels que la discipline, la participation, les travaux de groupe...

En somme, nous pouvons affirmer que les trois formes explicitées ci-dessus se complètent pour assurer des apprentissages réels qui seraient investis dans les situations nouvelles. Ce ne sont cependant des formes récentes dans la mesure où les formes formative et sommative étaient appliquées dans les pédagogies précédant l'application de l'APC. Néanmoins, cette nouvelle approche a introduit la forme diagnostique, une forme récente, et lui a accordé un statut indéniable. Étant donné que l'évaluation sommative est celle sur laquelle porte notre recherche, nous tentons de présenter dans l'élément suivant les outils évaluatifs préconisés par les instructions officielles.

### 3.2. Quels outils évaluatifs ?

Les documents consultés qui datent de 2010, 2013 et 2015 sont unanimes sur les outils évaluatifs à introduire dans le cadre de l'APC : les situations d'évaluation et les grilles critériées.

S'agissant des situations, elles constituent les outils de mesure préconisés par les instructions officielles, conformément aux principes de la nouvelle approche. C'est ce que le Programme de français de 2AM qui explicite : « *mettre l'élève en situation d'évaluation c'est le placer en situation d'intégration où il doit résoudre une situation problème en mobilisant un ensemble ressources* » (2010, p. 15). Cette précision, qui consolide des informations citées dans la partie théorique de notre étude, révèle la nécessité de recourir à des situations complexes en évaluation des compétences. Le « *Document d'accompagnement au programme de 3AM* » (2012-2013), qui se réfère aux travaux de Roegiers (2000, 2006), préconise les situations d'intégration en évaluation comme des instruments de mesure.

Si nous citons Roegiers à ce niveau de notre étude, c'est en raison de ces travaux sur l'évaluation des apprentissages dans l'ouvrage collectif « *La refonte de la pédagogie en Algérie- Défis et enjeux d'une société en mutation* » paru en 2005. Selon cet auteur, la nouvelle approche évalue les apprentissages à travers des situations complexes qui ont certaines caractéristiques. Il ne s'agit pas d'une évaluation de savoirs et de savoir-faire telle que la pédagogie précédente le suppose. Nous ne devons pas oublier que d'autres auteurs tels que Gérard (2006) ont fourni des précisions précieuses sur l'évaluation des compétences et sur les situations évaluatives. Ces précisions, nous en retrouvons certaines dans les documents consultés. En vue d'illustrer nos propos, nous citons le « *Document d'accompagnement du programme de français de 1AM* » (2010) qui liste les caractéristiques essentielles des situations. Mentionnons que des exemples de situations d'intégration sont proposées pour que les enseignants d'y appuient lors de la conception des situations d'intégration et d'évaluation. Voici

un exemple d'une situation d'intégration proposée dans le « *Document d'accompagnement au programme de français- 2<sup>ème</sup> année moyenne* ».

<b>Compétence</b> : « Produire un récit de fiction. »			
Situation d'intégration	Valeurs	Compétences transversales	Critères d'évaluation
<p>A l'occasion du 16 avril, journée du savoir, votre professeur propose à la classe de participer au concours régional du meilleur conte. Ce concours destiné aux jeunes écrivains comporte les conditions suivantes :</p> <p><b>Rédiger</b> un récit de fiction de plus de 150 mots.</p> <p><b>Traiter</b> un sujet sur le thème de la solidarité.</p> <p><b>Appliquer</b> le schéma actanciel.</p> <p><b>Faire</b> preuve d'imagination en illustrant le conte.</p> <p><b>Accompagner</b> le texte produit de son résumé.</p> <p><b>Présenter</b> le texte saisi à l'ordinateur.</p> <p>Ce travail réalisé avec l'ensemble de la classe fera l'objet d'une distribution des tâches. La production finale sera soumise à un jury de professeurs pour retenir les 10 meilleurs contes de la région.</p>	<p>Le choix du thème de la <b>solidarité</b> est déjà porteur de sens.</p> <p>Le travail est une production <b>collective</b> qui mettra à l'épreuve les élèves sur le plan de l'entraide, de l'échange, de la concertation en vue de la meilleure production.</p>	<p><b>Ordre intellectuel</b> :</p> <p>Résoudre des situations problèmes.</p> <p>Rechercher seul l'information utile pour résoudre le problème (utilisation d'internet et des bibliothèques).</p> <p><b>Ordre méthodologique</b> :</p> <p>Prendre des notes et les organiser.</p> <p>Synthétiser une information et en rendre compte sous une forme résumée par écrit.</p> <p>[...]</p> <p><b>Ordre de la communication</b> :</p> <p>Communication de façon intelligible, claire, précise et appropriée.</p> <p><b>Ordre personnel et social</b> :</p> <p>Interagir positivement en affirmant sa personnalité mais aussi en respectant l'avis des autres.</p> <p>S'intégrer à un travail d'équipe, un projet...</p>	<p><b>Pertinence de la production</b> :</p> <p>Utilisation des ressources proposées.</p> <p>Respect de la consigne.</p> <p><b>Correction de la langue</b> :</p> <p>Bonne orthographe des mots.</p> <p><b>Cohérence sémantique</b> :</p> <p>Respect de la structure du texte à produire.</p> <p>Application du schéma actanciel.</p> <p>Utilisation des temps verbaux adéquats.</p> <p>Richesse du vocabulaire.</p>
<b>Ressources et documents</b> : des contes, des nouvelles, un ordinateur, le schéma actanciel... »			

**Tableau 2: Exemple d'une situation d'intégration**  
(Document d'accompagnement au programme de français- 2<sup>ème</sup> année moyenne, 2011-2012, p. 44).

Nous admettons que cette situation qui présente le projet que les apprenants doivent réaliser est loin d'être une situation idéale dans la mesure où elle ne respecte pas toutes les conditions nécessaires. Néanmoins, nous ne pouvons pas nier la précision de plusieurs éléments tels que le contexte, le produit attendu, le but de la production, les critères et indicateurs d'évaluation...

En exposant cet aperçu sur l'évaluation dans les documents officiels, nous explicitons les instructions officielles préconisées pour une application rigoureuse des principes de l'APC, une précision très utile pour notre étude. Il est connu que notre étude porte sur la validité des épreuves évaluatives du FLE au cycle moyen qui suppose une certaine conformité aux exigences institutionnelles. Les épreuves que nous analysons dans notre deuxième chapitre pratique deviennent ainsi non valides si elles ne suivent pas toutes les exigences exposées ci-dessus. En effet, une évaluation de compétences recourt à des situations complexes permettant aux apprenants de manifester leurs compétences effectives.

Nous devons mentionner une certaine confusion terminologique qui ne passe pas inaperçue dans les documents consultés. Aucune distinction entre les notions de *situation d'intégration* et *situation d'évaluation* n'est ressentie dans les explications fournies. Cette confusion n'est pas certainement due à une ignorance terminologique dans la mesure où elle peut être aisément justifiée. Dans la définition des situations évaluatives, nous précisons toujours qu'elles sont des situations d'intégration, d'où le rapport de synonymie établie entre les notions en question. Néanmoins, ce rapport ne doit pas être adopté étant donné que la distinction entre les deux notions s'impose en raison de leurs objectifs distincts : la situation d'intégration vise un apprentissage de l'intégration alors que la situation d'évaluation vise la vérification de la maîtrise des apprentissages.

Ces situations évaluatives investies pour une évaluation des compétences des apprenants aboutissent à des productions que les enseignants doivent corriger en recourant à des outils précis. Quels sont les outils que les programmes et les documents officiels préconisent-ils ?

Les grilles critériées sont les outils de jugement préconisés pour évaluer les productions des apprenants. Cependant, dans les documents consultés, elles se présentent notamment comme des outils d'auto-évaluation et de co-évaluation, ce qui n'en occulte pas l'utilité en évaluation sommative. En tant qu'outil d'auto-évaluation, ces grilles permettent aux apprenants de réaliser à quel degré les apprentissages programmés sont maîtrisés, suscitant une réflexion nécessaire au développement des compétences.

Le « *Document d'accompagnement au programme de 3AM* » fournit des précisions et des exemples sur les grilles critériées. Selon ses concepteurs, ces grilles sont constituées de critères qui permettent de « *favoriser une centration sur la tâche à réaliser, faire prendre conscience des stratégies à utiliser, permettre la réflexion critique sur l'activité* » (2012-2013, p. 60). En effet, les critères retenus aux apprenants les aident à mieux réaliser les tâches car ils déterminent les éléments à prendre en considération. Aussi, ils attirent leur attention sur les stratégies à investir pour s'assurer de la qualité de leurs productions. Enfin, les critères

développent une certaine réflexivité permettant à chacun des apprenants d'adopter un regard critique sur son propre apprentissage. En vue de concevoir ces grilles destinées notamment à un usage formatif, les auteurs du document cité préconisent une construction collective (réunissant enseignants et apprenants) pour que les apprenants puissent progressivement élaborer des grilles individuelles pour s'évaluer.

Certes, ces précisions concernent la forme formative. Cependant, elles sont identiques à celles que les enseignants investissent pour une évaluation sommative des apprentissages. La grille suivante que le « *Document d'accompagnement au programme de 3AM* » propose pourrait être investie à cette fin. Cette grille, ci-dessous présentée, est constituée de trois critères (la pertinence, l'organisation et l'utilisation du code linguistique) qui se subdivisent en indicateurs explicitant chacun des critères cités.

Critères	Indicateurs
Pertinence	<ul style="list-style-type: none"> <li>- Respect de la grille proposée.</li> <li>- Respect du repère temporel : emploi du passé composé/plus-que-parfait.</li> <li>- Pertinence du titre.</li> </ul>
Organisation	<ul style="list-style-type: none"> <li>- Organisation chronologique avec emploi des indicateurs de temps.</li> <li>- Présence des reprises anaphoriques.</li> </ul>
Utilisation du code linguistique	<ul style="list-style-type: none"> <li>- Emploi correct du passé composé/imparfait/plus-que-parfait.</li> <li>- Présence des verbes introducteurs pour introduire des témoignages.</li> <li>- Lexique adéquat (insolite/faits divers).</li> <li>- Ponctuation correcte.</li> </ul>

**Tableau 3: Grille de l'enseignant**  
(Document d'accompagnement au programme de 3AM, 2012-2013, p. 61).

En dépit des limites de la grille proposée (redondance de certains indicateurs, incohérence entre critères et indicateurs...), nous trouvons que cette grille informe sur les instruments de jugement préconisés par les instructions officielles qui tentent réellement de respecter certaines exigences de l'APC. Les problèmes mentionnés au niveau de cette première grille semblent évités dans les exemples que contiennent les plans annuels parus en 2019. En effet, les critères et les indicateurs - qui y sont proposés- sont les suivants :

- Pertinence de la production : parmi les indicateurs qui révèlent la maîtrise de ce critère, les plans annuels des quatre niveaux du cycle moyen citent le respect de la consigne proposée.
- Correction de la langue : le respect de la structure de la phrase et celui des règles morphosyntaxiques figurent parmi les indicateurs énumérés pour évaluer la maîtrise de ce critère.

- Cohérence sémantique : la maîtrise de ce dernier critère est évaluée à travers le respect de la structure du type textuel à produire, la progression chronologique, l'emploi des temps verbaux adéquats...<sup>104</sup>

Les exemples que nous venons de citer révèlent un certain progrès quant au respect des exigences de l'APC dans les documents officiels étant donné que seuls trois critères minimaux sont cités, auxquels s'ajoute un critère de perfectionnement. Mais, aucun exemple des critères de perfectionnement n'est proposé par les concepteurs des plans annuels de 2019, ce qui laisse aux enseignants la tâche de les sélectionner. Mentionnons que les critères et indicateurs énumérés ne sont que des propositions et que les évaluateurs ne sont pas obligés de les intégrer tels que les plans annuels les exposent.

Nous pouvons ainsi affirmer que les instructions officielles que contiennent les documents, programmes...sont assez conformes aux exigences de l'APC dans la mesure où les outils de mesure (les situations d'évaluation) et de jugement (les grilles critériées) préconisés sont ceux prônés par cette nouvelle approche. Ces instructions clairement annoncées dans les documents qui datent de 2010 signifient que les évaluations sommatives proposées ces dernières années s'appuient sur les outils précédemment cités. C'est ce que nous vérifions à travers l'analyse de notre corpus que nous exposons dans les éléments qui suivent.

Consulté récemment, le « *Document d'accompagnement du programme de français-cycle moyen-*» (2016) propose un exemple d'une évaluation certificative<sup>105</sup> que les enseignants pourraient sélectionner pour évaluer leurs apprenants de 1AM. En dépit de quelques limites (au niveau du nombre de tâches et de la clarté des consignes), cette évaluation –citée ci-dessous– correspond à la description théorique des situations évaluatives et répond à de nombreuses exigences de l'APC.

**Situation**

Chaque année, des enfants piétons sont victimes d'accidents sur la route. Une campagne de sensibilisation aux abords des écoles est organisée par ta commune à l'aide d'affiches et de panneaux. Mais l'affiche de prévention accrochée près de ton collègue a été déchirée par le vent. Ta classe est alors sollicitée pour remplacer cette affiche par des prospectus à distribuer. Pour cela, ton prospectus comportera deux parties :

- dans la première partie, tu répondras à la question « Pourquoi y a-t-il autant d'accidents sur nos routes ? » ;

<sup>104</sup> Les trois critères minimaux sont cités tels que les plans annuels les proposent pour les différents niveaux en les adaptant selon les besoins. S'agissant des critères de perfectionnement, aucun critère n'est clairement mentionné.

<sup>105</sup> L'évaluation certificative désigne, dans ce document, les évaluations sommatives (trimestrielles et annuelles).

- dans la deuxième partie tu proposeras deux recommandations adressées aux automobilistes pour leur montrer les attitudes à suivre et trois recommandations à destination des enfants pour qu'ils évitent les comportements négatifs en tant que piétons.

**Ressources proposées :** Extrait de la Loi relative à l'organisation, la sécurité et la police de la sécurité routière ; Panneaux du code de la route ; Statistiques relatives aux principales causes des accidents ; Article « Le programme national «Permis piéton pour enfants»».

### Documents ressources

#### Section 1 : Définitions

- **trottoir**: un espace aménagé sur les côtés d'une route destiné à la circulation des piétons ; il doit être plus élevé que la chaussée et généralement bitumé ou dallé ;

- **conducteur**: toute personne qui assure la direction d'un véhicule, y compris les cycles et cyclomoteurs, guide d'animaux de trait, de charge, de selle, de troupeaux sur une route ou qui en a la maîtrise effective ;

- **piéton**: la personne se déplaçant à pied ;

- **automobile**: tout véhicule destiné au transport de personnes ou de marchandises et pourvu d'un dispositif mécanique de propulsion circulant sur route. (...).

#### Section 2 : Principes généraux

**Art. 9.** - Tout conducteur de véhicule doit observer les règles de la circulation routière de manière à ne constituer aucun danger pour lui-même et pour les autres usagers.

**Art. 23.** - Tout conducteur doit adapter la vitesse de son véhicule aux difficultés et obstacles de la circulation, à l'état de la chaussée et aux conditions météorologiques. Il doit constamment rester maître de la vitesse de son véhicule et conduire avec prudence.

**Art. 34.** - Les piétons sont tenus d'emprunter les trottoirs ou accotements spécialement aménagés à leur usage. Est interdite toute utilisation des trottoirs à des fins entravant la circulation piétonne.

**Art. 35.** - Les piétons sont tenus, pour traverser une chaussée, de s'assurer au préalable qu'il n'existe pas de danger immédiat et tenir compte également de la distance et de la vitesse des véhicules y circulant et d'utiliser les passages matérialisés, spécialement prévus à leur intention dits "passage piétons" toutes les fois qu'un tel passage se trouve à moins de 30 m. La traversée de la chaussée doit se faire en ligne droite, c'est à dire perpendiculairement, à l'axe de la chaussée. Il est interdit aux piétons de s'immobiliser sur la chaussée.

Extrait de La Loi n° 01-14 du 19 /08/ 2001 relative à l'organisation, la sécurité et la police de la circulation routière modifiée et complétée par l'ordonnance n°09-03 du 22/07/ 2009

#### Des permis piéton pour les enfants

Trois écoles primaires ont été retenues pour abriter des journées de sensibilisation sur la sécurité routière dans le cadre du programme national «*Permis piéton pour enfants*», initiées par la Gendarmerie nationale en collaboration avec le ministère de l'Education nationale.

Des séances d'apprentissage des règles de la circulation piétonne de 45 minutes seront dispensées aux élèves de la 5e année, chaque mardi. Des jeux de questions-réponses, des démonstrations et des vidéos leur seront proposés. Ce programme, qui prendra fin le 16 avril avec la distribution de permis au profit des enfants participants, a pour but de leur inculquer le sens de la responsabilité individuelle, en mettant en exergue un ensemble de précautions et gestes leur permettant d'assurer leur propre sécurité.





*N. Hammoud, ElWatan, 20/02/2016*

#### Les principales causes des accidents corporels durant la période allant du 08 au 14/03/2016

La cause	Le nombre	Le Pourcentage
Le facteur humain	156	88.63%
Le véhicule	13	07.39%
La route et l'environnement	07	03.98%
Total	176	100%

*Source site DGSN*

#### Des panneaux du code de la route

			
Passage pour piétons	Annonce de feux tricolores	Endroit fréquenté par les enfants	Débouché de cyclistes

**Tableau 4 : Exemple d'une épreuve d'évaluation (Document d'accompagnement du programme de français-cycle moyen, 2016, p. 93-94).**



Cette situation évaluative, qui vise une **seule production évaluable** (un texte constitué d'une partie explicative et d'une autre prescriptive), contient en réalité *trois tâches interdépendantes* : la règle des 2/3 y serait ainsi respectée si les consignes étaient plus claires et précises. Cette intention de proposer une **seule tâche** est plus ostensible dans la grille de correction que nous reprenons ci-dessous.

	Critères et notes / 20	Indicateurs	Absence de maîtrise	Maitrise partielle	Maitrise minimale	Maitrise totale
			Aucun indicateur	Présence d'un indicateur	Présence de deux indicateurs	Présence de tous les indicateurs
La maîtrise des ressources	Pertinence de la production ...points	-Utilisation des ressources proposées. -Respect de la consigne. -Utilisation de la typographie appropriée - Intégration de valeurs. - Respect de la mise en page du prospectus. - Utilisation de la typographie appropriée.	00	...points	...points	...points
	Correction de la langue ...points	- Respect de l'organisation de la phrase. - Bonne orthographe des mots. -Emploi correct des temps.	00	...points	...points	...points
	Cohérence sémantique ...points	- Respect de la structure du texte à produire (texte explicatif et prescriptif). -Utilisation des ressources linguistiques adéquates. - Progression thématique à thème constant.	00	...points	...points	...points
	Critère de perfectionnement ...points	- Richesse du vocabulaire. - Emploi de la couleur rouge pour l'interdiction.	00	...points	...points	...points
La mobilisation des ressources et des compétences transversales	- Résoudre une situation problème - Se positionner en tant que scripteur (a développé des méthodes de travail individuelles)	- s'implique dans une tâche individuelle. - exploite les ressources de la communication - utilise un brouillon, - révisé sa production. - remet un travail propre,				

Attitudes et valeurs	L'élève est sensibilisé à un problème de société	- participe à la vie collective, - développe un discours citoyen, - délimite ce qui relève des devoirs des automobilistes et des piétons, - transmet les principes de respect de la vie humaine.				
----------------------	--	---	--	--	--	--

**Tableau 5 : Grille d'évaluation de l'épreuve certificative ((Document d'accompagnement du programme de français-cycle moyen, 2016, p. 96-97).**

Bien que cette grille soit conforme à certaines exigences de l'APC, elle présente quelques limites : elle évalue une seule tâche, ce qui ne reflète pas les compétences effectives des évalués ; elle comprend des critères minimaux et de perfectionnement sans préciser le taux de points à leur accorder par les évaluateurs ; elle intègre des critères liés aux compétences transversales et aux valeurs sans les noter... Nous ne nous étalons pas sur ces limites étant donné que cette grille ainsi que l'épreuve -mentionnée supra-sont citées pour montrer que les instructions officielles sont conformes à certaines exigences théoriques de l'APC ; ce qui suscite des questionnements sur les raisons de la stagnation supposée des pratiques évaluatives.

#### 4. Des précisions sur les corpus<sup>106</sup> analysés

Dans les éléments qui suivent, nous exposons les précisions que nous jugeons nécessaires sur les corpus analysés en vue de répondre à la problématique de notre recherche. Nous devons mentionner que nous retenons deux corpus dans notre étude dont la sélection est imposée par les démarches analytique et comparative adoptées pour collecter des éléments de réponses. Concernant la démarche analytique, elle porte sur les épreuves d'évaluation sommative (devoirs et compositions) de tous les niveaux du cycle moyen pour l'identification des pratiques évaluatives mises en œuvre, ce qui assure par la suite l'étude de leur conformité aux instructions officielles et aux exigences de l'APC. Ces épreuves constituent ainsi notre premier corpus. Quant au second, il est constitué des examens de BEM que nous analysons pour vérifier si les pratiques évaluatives ont réellement évolué depuis l'application de la nouvelle approche. Si notre second corpus est réduit aux épreuves de BEM<sup>107</sup>, c'est pour deux raisons

<sup>106</sup>Le terme corpus correspond à « l'ensemble des unités (textes, objets ou autres) avec lesquels nous travaillerons et qui seront véritablement utilisée pour nos résultats » (Mongeau, 2008, p. 91).

<sup>107</sup>Même si ces évaluations sont introduites au terme du cycle moyen, nous ne les prenons comme des évaluations sommatives tout comme les compositions et les devoirs. Contrairement à ces deux modes d'évaluation sommative, le BEM doit porter sur des apprentissages de tout un cycle de formation. Néanmoins, cet examen n'aboutit pas à

essentielles. La première est le caractère représentatif incontestable de ces examens qui, vu leur aspect officiel, constituent des épreuves modèles et sont censés être conformes aux instructions officielles. En analysant ces épreuves, nous obtenons des réponses satisfaisantes sur la validité des évaluations sommatives du cycle moyen. La seconde raison est de l'ordre de la faisabilité : nous ne pouvons en aucun cas collecter des devoirs et des compositions qui reflètent les pratiques évaluatives précédant l'APC, mais nous réussissons aisément à collecter quelques examens de BEM qui précèdent la mise en œuvre de cette approche. Après avoir présenté brièvement les deux corpus de notre étude, nous en fournissons ci-dessous toutes les précisions nécessaires.

#### **4.1. Corpus 1 : devoirs et compositions**

Le premier corpus sur lequel porte notre analyse est constitué des devoirs et des compositions des trois trimestres<sup>108</sup>. Ces sujets évaluent le degré de maîtrise des apprentissages dispensés au terme des séquences et des projets. Ce point nous conduit à aborder le premier critère fondamental de la constitution de notre corpus qui regroupe toutes les épreuves évaluant les compétences des apprenants au terme d'une série d'apprentissages, c'est-à-dire les évaluations sommatives. Cependant, d'autres éléments, que nous énumérons ci-dessous, déterminent davantage notre corpus. Il s'agit de :

##### **4.1.1. Taille du corpus**

Il est connu que le cycle moyen est constitué de quatre niveaux. Pour obtenir un corpus suffisamment représentatif des évaluations sommatives écrites introduites, une centaine (100) de sujets est retenue pour chaque niveau. Au total, quatre cents (400) devoirs et compositions sont analysés dans notre étude pour identifier les pratiques évaluatives adoptées par les enseignants du cycle. Mentionnons que la distinction entre devoirs et compositions, qui n'est pas mise en valeur dans cette étude, est peu pertinente étant donné que les deux évaluations sommatives adoptent exactement la même démarche, d'où notre volonté de les étudier indissociablement.

Le nombre de sujets à analyser fixé ne correspond pas à celui des sujets collectés. Ainsi, pour la première année moyenne, nous avons collecté cent-quarante-trois (143) sujets.

---

l'obtention d'une attestation ou un certificat validant les compétences des apprenants en FLE qui ne constitue qu'une discipline parmi les autres. Mentionnons que les examens de BEM sont repris par certains enseignants en tant que devoirs ou compositions.

<sup>108</sup> Les contenus de chaque année scolaire sont répartis sur trois trimestres. Au cours de chaque trimestre, qui s'étale sur environ trois mois, deux devoirs (mais parfois un seul) et une composition évaluent les apprentissages des apprenants.

S'agissant de la deuxième année moyenne, cent-trente-sept (137) sujets sont collectés. Le nombre des sujets de deuxième année collectés est plus élevé que celui de la troisième année dont le nombre est cent-huit (108) sujets. C'est en raison de ce nombre-ci que le nombre de sujets analysés pour chaque niveau était fixé à cent (100) pour une certaine homogénéisation entre les niveaux. Enfin, le nombre des sujets de quatrième année collectés, le plus élevé de tous les niveaux, est cent-cinquante-huit (158).

De ces sujets collectés dans les différents niveaux, plusieurs ont été écartés de l'étude pour des raisons variées. Ces raisons, nous les classons dans trois grandes catégories. La première englobe les sujets qui ne contiennent pas tous les renseignements retenus dans notre grille d'analyse : le type de l'épreuve, l'année scolaire, le trimestre, la durée... Quelques renseignements, tels que l'année scolaire, sont en effet indispensables dans la mesure où ils sont décisifs dans la constitution du corpus à analyser. Ainsi, le manque de précisions sur les sujets a écarté douze (12) sujets de première année et six (06) sujets de deuxième année. La deuxième catégorie est celle des sujets qui présentent des problèmes de lisibilité qu'ils soient imprimés ou électroniques. Nous les écartons pour éviter certains biais qui pourraient fausser les résultats de notre étude (une mauvaise qualité d'impression, des questions incomplètes...). C'est pour cette raison d'ailleurs que quelques sujets de quatrième année ont été omis de notre corpus. La dernière catégorie de sujets nous paraît plus importante que les deux précédentes étant donné qu'elle dévoile une tradition enseignante perpétuée dans l'univers scolaire. Ce sont les sujets qui se répètent : le devoir d'un établissement servira de composition d'un autre, la composition qu'un enseignant introduit durant le premier trimestre est celle qu'un autre introduit dans le deuxième semestre, un sujet similaire proposé dans plusieurs établissements sans qu'il y ait une coordination entre les enseignants... Nous ne nous attardons pas sur les raisons qui amènent les enseignants à introduire des épreuves évaluatives qui ne correspondent pas toujours aux enseignements dispensés, un élément peu pertinent pour notre étude. Nous mentionnons surtout que quelques sujets identiques ont été écartés de notre corpus. Ce sont des sujets qui contiennent les mêmes textes, les mêmes questions de compréhension et de grammaire ainsi que les mêmes consignes de productions.

Après avoir écarté quelques sujets, nous avons retenu au total quatre cents (400) sujets, variés entre devoirs et compositions. Ce sont ces sujets qui nous permettent de repérer les pratiques auxquelles les enseignants recourent pour évaluer les compétences des apprenants en FLE. Cette taille de notre corpus, nous la jugeons assez représentative pour obtenir des résultats pertinents notamment que notre étude vise des données qualitatives sur les épreuves à analyser.

#### **4.1.2. Bornes chronologiques**

A travers cet élément, nous présentons un deuxième critère qui nous a permis la constitution de notre corpus : les années scolaires des sujets retenus. Nous ne pouvons sélectionner les sujets à analyser sans tenir compte des années scolaires de leur introduction vue la refonte pédagogique qu'a connue l'école algérienne. Il est connu que « *cette refonte a donné naissance à deux réformes successives : la réforme de la première génération de 2003 et de la deuxième génération de 2016* » (Bouamama & Boubakour, 2018, p. 130). La première, mentionnée ci-dessus, consiste au passage de la PPO à l'APC en 2003 qui a provoqué une rupture avec les pratiques enseignantes et évaluatives de l'ancienne pédagogie. Quant à la seconde, elle se présente comme une réécriture des programmes et notamment des manuels de la première génération pour combler les insuffisances constatées et renforcer le rôle des apprenants.

Ces deux réformes exigent la constitution de notre corpus à partir des épreuves introduites à l'ère de la deuxième génération étant donné que les manuels ainsi que les programmes de la première génération révisés ne reflètent plus suffisamment les exigences et des instructions institutionnelles. De ce fait, les épreuves collectées sont celles introduites durant les années scolaires suivantes : 2017/2018, 2018/2019 et 2019/2020. A première vue, il paraît plus pratique de restreindre notre analyse aux sujets les plus récents pour des résultats assez pertinents et valides. Cependant, il serait intéressant de vérifier si les pratiques évaluatives ont évolué pour répondre aux exigences, d'où le maintien de tous les sujets de la deuxième génération.

Nous devons signaler que les pratiques évaluatives qui s'inscrivent dans le cadre de la nouvelle approche sont censées maîtrisées par les enseignants du cycle moyen qui ont bénéficié de plusieurs formations assurées par les inspecteurs soucieux de communiquer aux enseignants toutes les nouveautés ainsi que les modalités de leur application. Outre les inspecteurs, des enseignants formateurs continuent à encadrer les enseignants nouvellement recrutés pour les préparer pleinement à la complexité du terrain pédagogique. Néanmoins, aucune formation-aussi fructueuse soit-elle- ne pourrait se substituer à l'auto-formation qui demeure la plus décisive si les enseignants veulent vraiment s'adapter aux changements méthodologiques adoptés par l'école algérienne.

#### **4.1.3. Bornes géographiques**

Notre étude porte sur la validité de l'évaluation sommative au cycle moyen en Algérie. Bien que notre corpus ne soit pas large au point de couvrir toutes les régions et tous les

établissements du territoire national, nous avons opté pour un contexte national au lieu d'un contexte régional pour plusieurs raisons.

La première est que les programmes, les contenus, les manuels, les modalités d'enseignement et d'évaluation ainsi que les instructions institutionnelles sont unifiés et destinés à tous les enseignants du cycle sans particularités régionales. Cette homogénéisation des prescriptions conduit à celle des pratiques enseignantes et évaluatives, d'où la similitude des épreuves conçues en vue d'évaluer les compétences des apprenants. Ces épreuves, en dépit de quelques différences liées notamment aux apprentissages évalués, obéissent à une structure identique. Il serait alors légitime de considérer que les épreuves collectées sont représentatives des pratiques réelles des enseignants des diverses régions du territoire national.

La seconde raison est que les épreuves que nous retrouvons dans les sites pédagogiques sont adaptées puis introduites par des enseignants exerçant dans des régions variées, étant donné que les pratiques au cycle moyen sont identiques. C'est pour cette raison que, parmi les sujets collectés, figurent des sujets identiques qui ont évalué des apprenants inscrits dans des établissements différents et parfois dans des régions diverses. Dans ce cas-ci, il serait injuste de réduire les résultats d'une étude nationale à une région particulière du pays.

Pour que notre étude soit nationale, nous ne devons pas collecter des épreuves sommatives de tous les établissements du cycle moyen qui existent dans toutes les régions du pays. Nous nous limitons à collecter des sujets assez variés pour représenter certaines pratiques évaluatives prônées par les enseignants pour en étudier la validité. Comment ces sujets variés sont-ils collectés ?

#### **4.1.4. Méthodes de collecte**

La dernière précision que nous fournissons sur notre corpus est les méthodes empruntées pour le constituer. Pour ce faire, nous exposons les deux catégories de sujets sur lesquelles porte notre analyse : les sujets imprimés et les sujets électroniques.

Les premiers sont récupérés dans des établissements variés situés dans différentes wilayas, notamment celles de l'Est algérien. En les récupérant, les enseignants-concepteurs nous ont expliqué certains éléments sur leur démarche dans la conception des sujets de devoirs et de compositions, des explications qui pourraient nous servir à comprendre certains résultats de notre étude.

Pour les seconds sujets, ils sont téléchargés de sites éducatifs variés. Le site le plus important qui nous a servi de référence est Ency Éducation<sup>109</sup>, un site pédagogique destiné tant aux enseignants qu'aux apprenants. Les enseignants y trouvent des documents variés comme des fiches pédagogiques, des livres électroniques, des programmes de toutes les disciplines, des plans annuels, des épreuves évaluatives...qui les aident à améliorer leurs pratiques. Quant aux apprenants, le site leur offre tous les documents dont ils ont besoin pour un meilleur rendement scolaire : des cours, des résumés, des exercices corrigés, des propositions de devoirs et de compositions... Ce site ne s'adresse pas à une catégorie précise d'apprenants : les écoliers, les collégiens, les lycéens et même les universitaires peuvent accéder aux contenus publiés qui répondent à leurs besoins.

La richesse des ressources, que ce site met à la disposition de tous les acteurs (enseignants, apprenants, parents...), n'est pas l'unique raison unique qui nous encourage à en extraire certains sujets que nous analyserons dans la suite de notre travail. C'est notamment la fidélité et la crédibilité des contenus qui y sont publiés. Contrairement à d'autres sites éducatifs où les contributions sont ouvertes à tous les utilisateurs désirant les enrichir, les gestionnaires de ce site procèdent à une vérification des informations et des contenus envoyés, en respectant certaines conditions, pour les valider.

Pour récapituler ce qui vient d'être exposé, notre premier corpus que nous analysons est constitué des épreuves écrites (devoirs et de compositions) de la deuxième génération, introduites au cycle moyen en Algérie pour évaluer les compétences des apprenants en FLE.

#### **4.2. Corpus 2 : les examens de BEM**

Tout comme le premier corpus, ce second corpus sera également analysé. Cependant, contrairement au premier dont les résultats dressent le bilan des pratiques introduites par les enseignants pour évaluer les compétences des apprenants en FLE, ce second corpus est investi dans une étude comparative qui confronte les examens de BEM introduits avant l'application de l'APC aux examens introduits suite à son adoption. Cette confrontation permet de déceler les changements que l'évaluation du FLE au cycle moyen doit à la nouvelle approche. Dans ce qui suit, nous fournissons quelques précisions sur ce second corpus.

---

<sup>109</sup> Lien vers le site Ency Education : <https://www.ency-education.com/>

#### 4.2.1. Taille

Notre corpus est constitué des examens de BEM qui évaluent le FLE. Mentionnons que le BEM est un examen national que passent tous les apprenants de la quatrième année moyenne, ce qui renforce la représentativité de notre corpus. Celui-ci comprend vingt (20) sujets dont la structure est identique à celle des épreuves sommatives ci-dessus abordés. Ce qui ces examens particuliers est leur introduction au bout d'un cycle : le cycle moyen se solde par ces examens que les apprenants doivent réussir pour pouvoir passer au lycée.

Les examens retenus dans ce corpus ne relèvent pas de la même pédagogie. Dix (10) examens retenus étaient introduits avant l'application de l'APC à l'enseignement/apprentissage de toutes les disciplines, y compris le FLE ; alors que les dix (10) sujets restants étaient introduits à l'ère de la présente approche.

#### 4.2.2. Précisions chronologiques

Nous avons pu collecter tous les examens de BEM introduits de 1996 à 2019. Notons que de 1996 à 1998, un seul sujet est introduit pour évaluer les apprentissages en FLE. Cependant, de 1999 à 2007, nous assistons à l'introduction de deux sujets en évaluation du FLE : le premier évalue le français en tant que première langue étrangère (LE 1) alors que le second l'évalue comme une deuxième langue étrangère (LE 2). Cette situation ne dure pas longtemps étant donné qu'un retour vers la situation initiale a lieu en 2008.

Cet exposé succinct révèle le recours à deux sujets en évaluation du FLE durant une certaine période qui s'est caractérisée par la complexité de la situation linguistique régnant le pays. En effet, la langue anglaise s'est imposée en 1993 comme première langue étrangère en occurrence avec le français. Dès la quatrième année primaire, les parents choisissent la première langue étrangère à apprendre par leurs enfants (Baala-Boudebia, 2012). Expérimentée de 1994 à 1998 puis généralisée, cette mesure explique le recours à deux sujets pour évaluer les deux langues étrangères présentes dans le milieu scolaire. Ainsi, les apprenants qui avaient appris la langue anglaise comme première langue étrangère sont évalués par des sujets où « *anglais première langue étrangère* » et « *français deuxième langue étrangère* » sont mentionnés. Ceux qui ont appris le français depuis la quatrième année primaire passent leur examen BEM en « *français première langue étrangère* » et « *anglais deuxième langue étrangère* ». Cette situation ne dure pas étant donné que la majorité des parents étaient défavorables à l'apprentissage précoce de l'anglais, selon des enquêtes menées par le ministère de l'Éducation nationale (Baala-Boudebia, 2012 ; Derradji, 2002).



Vu que deux sujets ont existé pour évaluer les apprenants algériens en FLE, il devient nécessaire de préciser lesquels intègre notre corpus. Étant donné que notre étude confronte les sujets post-APC<sup>110</sup> aux sujets pré-APC<sup>111</sup> et que les sujets post-APC évaluent la langue française, en tant que première langue étrangère, nous ne retenons que les épreuves qui évaluent le français appris dès la quatrième année primaire, le français en tant que première langue étrangère. Les sujets proposés de 1997 à 2006 représentent les pratiques évaluatives qui précèdent l'APC alors que ceux proposés de 2010 à 2019 reflètent les pratiques qui lui sont postérieures.

De prime abord, il semble incohérent de ne pas fixer 2003 comme une année transitoire car cette année-là a connu l'introduction de la nouvelle approche en Algérie. Cependant, en prenant en considération que le BEM est destiné aux apprenants de la quatrième année moyenne, l'année 2007 devient la première année où les examens évaluant le FLE, ainsi que les autres disciplines, s'inscrivent dans le cadre de l'APC. C'est pour cette raison que les sujets pré-APC s'étendent à 2006. Quant aux sujets post-APC, nous n'avons pas intégré les années 2007, 2008 et 2009 dont les examens doivent intégrer progressivement les prescriptions de la nouvelle approche qui serait suffisamment suivie depuis 2010.

#### **4.2.3. Méthodes de collecte**

En vue de collecter les examens de BEM sur lesquels porte notre étude, nous avons consulté le même site précédemment cité (Ency Éducation), qui regroupe tous les sujets proposés pour évaluer le FLE de 1996 à 2019. Ces sujets sont accompagnés des corrigés-types et barèmes qui guident la correction que les enseignants-correcteurs doivent effectuer. Ces corrigés-types et barèmes, nous les étudierons brièvement pour recueillir des données sur les outils de jugement investis.

En somme, le recours à deux corpus dans notre travail est dicté par les deux approches empruntées pour vérifier la validité des dispositifs d'évaluation sommative du FLE actuellement mis en œuvre. Le premier est étudié pour vérifier cette validité tandis que le second est intégré pour renforcer les résultats obtenus à travers l'analyse du premier corpus. Cette variation des méthodes et des corpus traduit notre souci de parvenir à travers la présente étude à des résultats assez fondés, qui lui procurent une certaine scientificité.

---

<sup>110</sup> Sujets introduits après l'application de l'APC.

<sup>111</sup> Sujets introduits avant l'introduction de l'APC en Algérie.

## Conclusion

A travers ce premier chapitre, nous avons explicité le cadre méthodologique de notre étude. Nous l'avons entamé par une explication détaillée des approches adoptées pour la collecte et l'analyse des données. Tenant compte des objectifs de notre étude, nous avons emprunté une approche qualitative qui décrit les pratiques évaluatives actuelles pour déterminer leur conformité aux exigences institutionnelles dictées par l'APC ; ainsi que la démarche quantitative pour déterminer à quel degré les pratiques nouvelles sont introduites par les enseignants du FLE exerçant au cycle moyen. En privilégiant cette démarche mixte, nous obtenons des données quantitatives et qualitatives qui nous permettent de comprendre pleinement l'évaluation des compétences dans le système éducatif algérien qui constitue le terrain de notre étude, avant de généraliser les résultats obtenus aux autres cycles étant donné que les pratiques évaluatives du cycle moyen sont le miroir qui nous reflète l'image de l'évaluation dans tous les cycles de l'Éducation nationale. Certes, nous ne prétendons pas que les résultats auxquels nous parviendrons seront généralisables sans réserves ; mais vu l'échantillon retenu, elles pourraient donner un aperçu sur les pratiques évaluatives de tous les cycles.

L'échantillon retenu est donc les épreuves introduites au cycle moyen pour une évaluation sommative des apprenants en FLE. Deux corpus sont retenus à cet effet pour pouvoir répondre à nos questions de recherches ; puis analysés séparément. Le premier est constitué des devoirs et compositions introduits au cours des années scolaires 2017/2018, 2018/2019, 2019/2020 ; c'est-à-dire après la réforme de la deuxième génération qui a apporté des changements qualitatifs profonds aux programmes, contenus, manuels et pratiques mise en œuvre suite à la première phase de réforme entamée en 2003, qui a opté pour l'APC. En analysant ces épreuves, récentes, de la deuxième génération qui doivent être les plus adaptées aux nouvelles exigences, nous pouvons étudier la validité de l'évaluation sommative au cycle moyen. A ce niveau, mentionnons que l'étude des documents et programmes du cycle moyen révèlent la préconisation des outils évaluatifs privilégiés par les recherches internationales dans le cadre de l'APC telles que les situations évaluatives et les grilles critériées. Si ces prescriptions sont proposées dans les documents de 2010, elles doivent être largement appliquées ces dernières années.

En vue de consolider les résultats de la première méthode analytique, la méthode comparative est empruntée. A travers celle-ci, nous confrontons des épreuves anciennes conçues avant l'intégration de la réforme de l'Éducation nationale aux épreuves conçues suite à la réforme. Néanmoins, des problèmes surgissent à ce niveau vu la difficulté de collecter des

épreuves qui remontent aux années précédant l'application de l'APC. Tenant compte de ces problèmes, nous avons limité notre analyse à des épreuves plus représentatives : ce sont les examens nationaux et officiels de BEM. Ainsi, nous comparons ceux intégrés avant 2007 à ceux introduits après cette date. A ce niveau également, certaines remarques se sont avérées indispensables ; car certaines réformes nationales de cette période doivent être évoquées. De toutes les années scolaires, la période qui s'étale de 1999 à 2007 a connu le recours à deux sujets pour évaluer le FLE. Cette particularité résulte de la décision institutionnelle qui a laissé aux parents la responsabilité de choisir la première langue à apprendre par leurs apprenants. C'est ainsi que le FLE est évalué en tant que première langue étrangère (LE1) ou en tant que deuxième langue étrangère (LE2) selon le statut accordé aux deux langues étrangères en présence : français et anglais. Seules les épreuves évaluant le FLE comme première langue étrangère sont retenues, car elles sont similaires au statut actuel de cette langue, pour que la comparaison soit légitime. En effet, celle-ci écarte une variable qui peut considérablement affecter la crédibilité des résultats de notre étude étant donné que les objectifs, les programmes, les contenus ainsi que les manuels retenus pour enseigner le français comme première langue étrangère sont complètement différents de ceux exploités pour l'enseigner en tant que deuxième langue étrangère. Cette confrontation des épreuves relevant de pédagogies différentes nous permettra également de vérifier si une évolution réelle au niveau des pratiques évaluatives existe dans le milieu scolaire, un élément qui pourrait renforcer les résultats obtenus à travers la méthode analytique.

Pour effectuer ces analyses et collecter des données utiles à notre étude, deux instruments (grille d'analyse et liste de vérification) sont investis. Une première grille est conçue pour une analyse des épreuves sommatives collectées. Nous l'avons constituée en nous référant à de nombreuses études nationales et internationales sur la validité. Certains indicateurs de validité sont certes plus décisifs que d'autres ; mais nous retenons tous les éléments pour parvenir à une description minutieuse de l'évaluation telle que les enseignants l'introduisent. Pour le deuxième outil ou la liste de vérification, elle servira à valider les épreuves conçues et proposées conformément aux exigences des programmes officiels et de l'APC.

En exposant toutes ces précisions, nous jugeons que le cadre méthodologique de notre étude est suffisamment précisé, ce qui nous permet de passer aux résultats obtenus, que nous exposerons et analyserons dans le chapitre suivant.

# **CHAPITRE 2 : RÉSULTATS, ANALYSE, ET DISCUSSION**

## Introduction

Le premier chapitre pratique était consacré au cadre méthodologique de notre recherche. De nombreuses précisions y étaient insérées sur les approches adoptées (analytique et comparative), les corpus à étudier, les outils de collecte... Ce deuxième chapitre exposera les résultats obtenus puis les analysera pour apporter quelques réponses à notre problématique : les évaluations sommatives du FLE sont-elles valides pour révéler les compétences réelles des collégiens algériens ? Ces résultats, défilés suite à la présentation des épreuves analysées, concernent notre premier corpus : les devoirs et les compositions des trois trimestres, conçus par des enseignants qui exercent dans divers établissements du territoire national. Rappelons que les épreuves sélectionnées étaient introduites dans les trois dernières années scolaires : 2017/2018, 2018/2019 et 2019/2020, ce qui révélera si -après 17 ans de la réforme du système éducatif- les pratiques évaluatives sont conformes aux instructions officielles.

Certaines précisions qui se rapportent à la présentation des résultats dans notre étude sont à mentionner. Ainsi, les résultats liés à certains indicateurs ne suivront pas leur ordre dans la grille conçue notamment lorsque des comparaisons sont jugées nécessaires pour une analyse pertinente. C'est le cas du nombre des items qui évaluent la compréhension et de celui des questions portant sur les ressources qui devront être présentés successivement puis ensemble pour déterminer si le statut accordé aux outils linguistiques en évaluation ne dépasse pas les exigences.

La deuxième précision concerne les résultats relatifs aux quatre niveaux du cycle moyen qui seront parfois présentés séparément vu les différences significatives qui affectent notre analyse (nombre des supports, leur nature...). Dans le cas échéant, les résultats des quatre niveaux seront présentés indissociablement. C'est le cas de certaines informations qui concernent les épreuves analysées (la durée, le type de l'épreuve...) ou précisent les compétences évaluées.

Les deux types d'épreuves analysées (devoirs et compositions) seront également présentés séparément lorsque nous présenterons les résultats liés à la longueur des supports textuels sélectionnés par les enseignants étant donné que des écarts importants ont été constatés.

Leur présentation aussi variée soit-elle, les résultats obtenus et leur analyse seront exposés en parallèle, ce qui permettra de vérifier si chacun des indicateurs exigés pour la validité des épreuves analysées y est respecté. Mentionnons que les indicateurs retenus dans notre grille conçue sont inspirés des travaux sur la validité des tests, évoqués dans notre volet théorique (De Ketele & Gérard, 2005 ; Roegiers, 2006 ; Gérard, 2009...).

Suite à la présentation des résultats et leur analyse, une discussion sera proposée à travers laquelle nous nous pencherons sur des études qui approuveraient ou rejetteraient les conclusions primaires formulées dans notre analyse. C'est à ce niveau de notre recherche que les résultats de la deuxième démarche comparative intégrée dans notre étude ainsi que leur analyse seront insérés. Nous procéderons en effet à la confrontation de deux catégories de BEM : ceux introduits avant la réforme (désignés BEM pré-APC dans notre étude) et ceux proposés après la réforme (dits BEM post-APC). Les conclusions que nous formulerons suite à cette comparaison appuieront les conclusions primaires obtenues grâce à la démarche analytique.

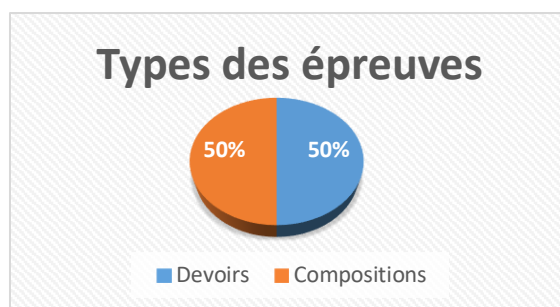
S'agissant des épreuves de BEM constituant notre second corpus, elles sont réparties en deux catégories : celles proposées dans les années *1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006* sont rangées dans la catégorie des BEM pré-APC, tandis que les BEM post-APC analysées sont celles proposées dans les années suivantes : *2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019*.

La comparaison effectuée exploite la même grille d'analyse conçue dans la démarche analytique : ses indicateurs permettent une description claire et précise des deux catégories de BEM. Cependant, certains indicateurs seront écartés pour des raisons dûment justifiées. Ainsi, la nouveauté des situations évaluatives conçues dans les BEM pré-APC ne pourrait être vérifiée sans le recours aux anciens manuels scolaires introduits au cours des années ayant précédé la réforme éducative. Vu la difficulté d'accéder à ces manuels pour vérifier si les situations d'écriture sélectionnées dans les BEM sont identiques à celles travaillées en classe, cet indicateur sera écarté, ce qui ne signifie pas que les indicateurs retenus ne seront pas suffisants pour atteindre nos objectifs.

Grâce aux deux méthodes investies, nous parviendrons à des conclusions suffisamment fondées et pertinentes qui pourraient contribuer à la compréhension de la réalité de l'évaluation scolaire du FLE dans notre contexte ; une compréhension indispensable à la résolution des problèmes éventuels.

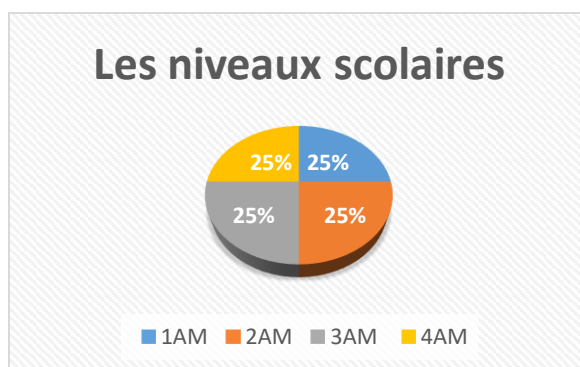
### **1. A propos de notre corpus**

Les épreuves sommatives analysées dans notre étude sont distinguées selon certaines variables que nous représentons dans les graphiques ci-dessous.



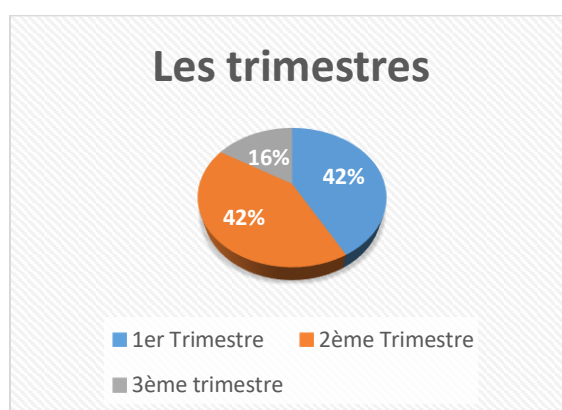
**Graphique 1: Répartition des types d'évaluations constituant notre corpus**

Ce premier graphique représente les deux types d'évaluations sommatives que notre étude analyse : devoirs et compositions dont le nombre est identique. Cette similitude du nombre est intentionnelle pour des explications plus pertinentes de certains éléments abordés. Mentionnons que ces épreuves relèvent des quatre niveaux du cycle moyen comme le représente le graphique suivant :



**Graphique 2: Répartition des niveaux du cycle moyen introduits dans l'étude**

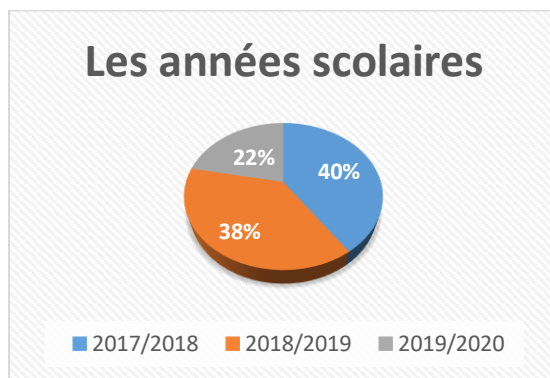
Les quatre niveaux du cycle moyen sont ainsi introduits dans notre étude, comme nous l'avons mentionné précédemment, à travers un nombre identique de devoirs et de compositions. Ces évaluations analysées couvrent tous les trimestres de l'année scolaire.



**Graphique 3: Répartition des trimestres scolaires couverts par notre corpus**

Il est clair que cette fois-ci le nombre des épreuves qui relèvent du troisième trimestre est nettement inférieur à celui des deux premiers, ce qui s'explique par plusieurs éléments. Lors de la constitution de notre corpus, nous n'étions pas soucieux de sélectionner un nombre

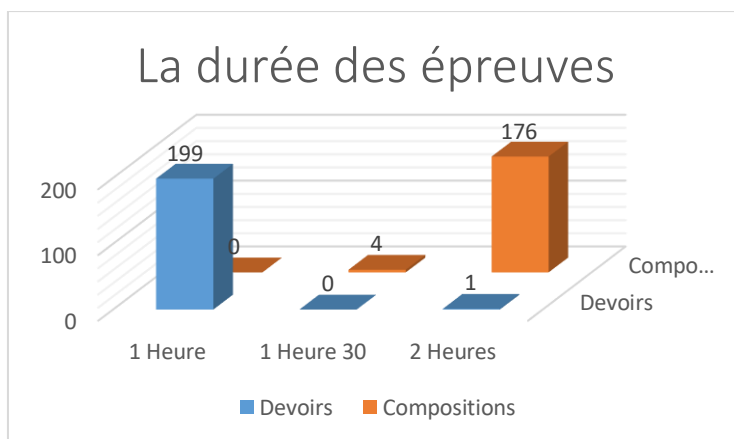
d'épreuves similaire dans les trois trimestres car cette distinction, contrairement aux deux précédentes, nous paraît peu significative étant donné que des frontières entre les contenus des divers trimestres ne pourraient jamais être établies. En effet, les enseignants sélectionnent les contenus à évaluer en fonction de leur état d'avancement qui est conditionné par de nombreux facteurs (personnels ou professionnels). Aussi, les enseignants sollicités ou contributeurs, dans les sites pédagogiques consultés, ont tendance à soumettre plus d'épreuves des deux premiers trimestres conçues plus soigneusement. Le troisième élément relève des circonstances particulières des dernières années (crise politique du pays et pandémie) qui ont eu un impact indéniable sur le déroulement des cours et des évaluations. S'agissant de l'année scolaire 2019/2020, le troisième trimestre, à maintes reprises reporté, était annulé. Dans ce qui suit, nous exposons la répartition de notre corpus sur les trois années scolaires retenues.



**Graphique 4: Répartition des années scolaires englobés par notre corpus**

Les épreuves de la présente année scolaire sont moins nombreuses que celles des années scolaires précédentes pour les raisons mentionnées supra.

La dernière précision représentée par le graphique ci-dessous concerne la durée des épreuves analysées.



**Graphique 5: Durée des épreuves constituant notre corpus**

Ce dernier graphique qui représente la durée des épreuves met en exergue une comparaison entre les temps impartis aux deux types d'évaluation sommative étudiés. En effet,



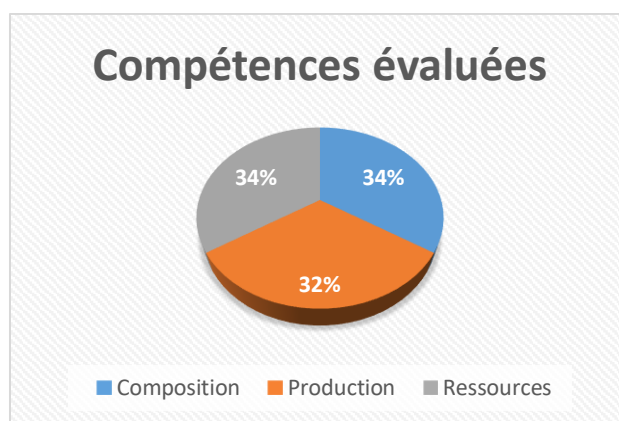
les enseignants accordent souvent une heure aux devoirs tandis que deux heures sont souvent imparties aux compositions, leur léguant par-là une certaine supériorité. Il est important cependant de vérifier si des différences significatives entre devoirs et compositions justifient la durée qui leur est accordée.

Suite aux précisions fournies sur notre corpus, nous passons dans ce qui suit à la présentation des résultats liés aux indicateurs que comporte notre grille. Nous les présentons, par catégories, en respectant leur ordre dans notre instrument de collecte.

## 2. Résultats et analyse

### 2.1. Concernant les compétences évaluées

Dans cette catégorie, nous présentons les trois premiers indicateurs indissociablement. Cependant, les trois derniers sont présentés par le même graphique pour des comparaisons entre les outils évaluant les diverses compétences. Alors, quelles compétences les sujets analysés évaluent-ils ? A cette question, répond le graphique ci-dessous :

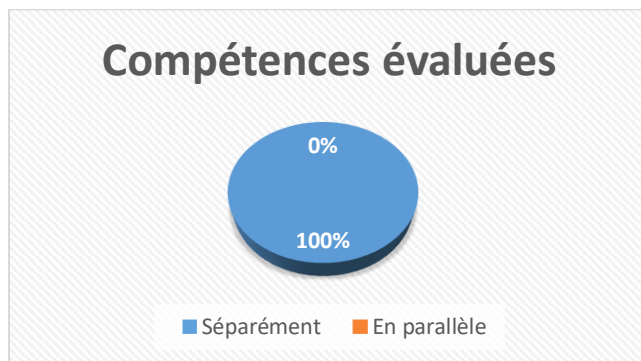


Graphique 6: Répartition des compétences évaluées

Trois composantes essentielles sont évaluées à travers les épreuves introduites au cycle moyen : la compréhension, la production et les ressources, qui renvoient aux points grammaticaux enseignés au cours des diverses années scolaires. Si l'évaluation des deux premières compétences, compréhension et production, semble tout-à-fait naturelle -étant donné que leur maîtrise est nécessaire pour communiquer en langue étrangère-, celle des ressources ne passe pas inaperçue dans la mesure où une évaluation sommative ne doit pas porter sur les ressources. Comme des auteurs tels que Gérard (2009) admettent une évaluation des ressources même dans une perspective sommative, il convient de vérifier le taux de points que les épreuves analysées leur consacrent pour des conclusions plus fondées sur la validité des épreuves étudiées.

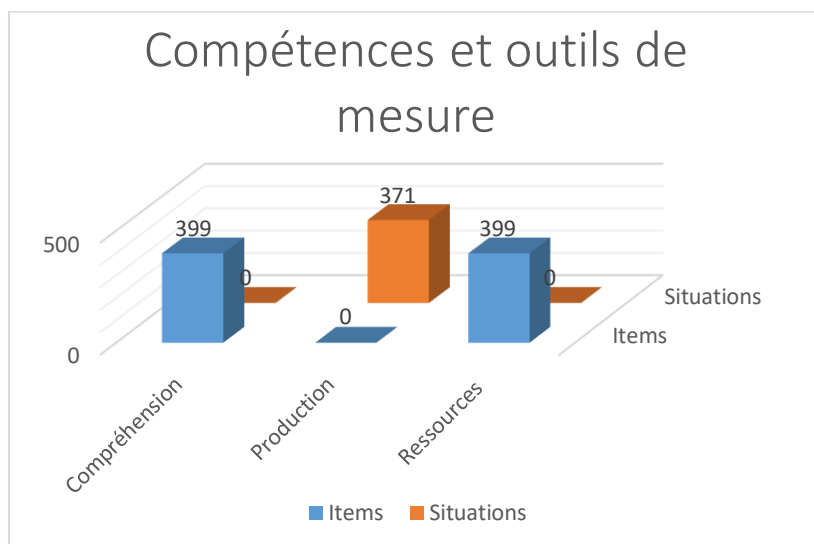
Les exigences en évaluation scolaire dans le cadre de l'APC ne concernent pas uniquement les ressources. Elles se rapportent également à la manière dont les deux

compétences clés sont évaluées dans les épreuves : ensemble ou séparément. En effet, ces deux compétences doivent être évaluées indissociablement à travers des situations mobilisant compréhension et production étant donné que les interactions linguistiques requièrent leur complémentarité. Dans cette optique, l'évaluation séparée des compétences est considérée classique par de nombreux auteurs tels que Huver et Spinger (2011). Alors, cette exigence est-elle respectée au cycle moyen ?



**Graphique 7: Répartition de la manière évaluant les deux compétences**

Ce graphique révèle que toutes les épreuves sommatives du FLE évaluent séparément les deux compétences, ce qui n'est pas conforme aux exigences, et suscite des interrogations sur la nature des outils évaluatifs introduits. Le graphique suivant répond à l'une de ces interrogations et montre que les enseignants recourent à des items (questions) pour évaluer la compréhension et les ressources (grammaire) alors que les situations sont privilégiées en évaluation de la production.



**Graphique 8: Outils de mesure évaluant les différentes compétences**

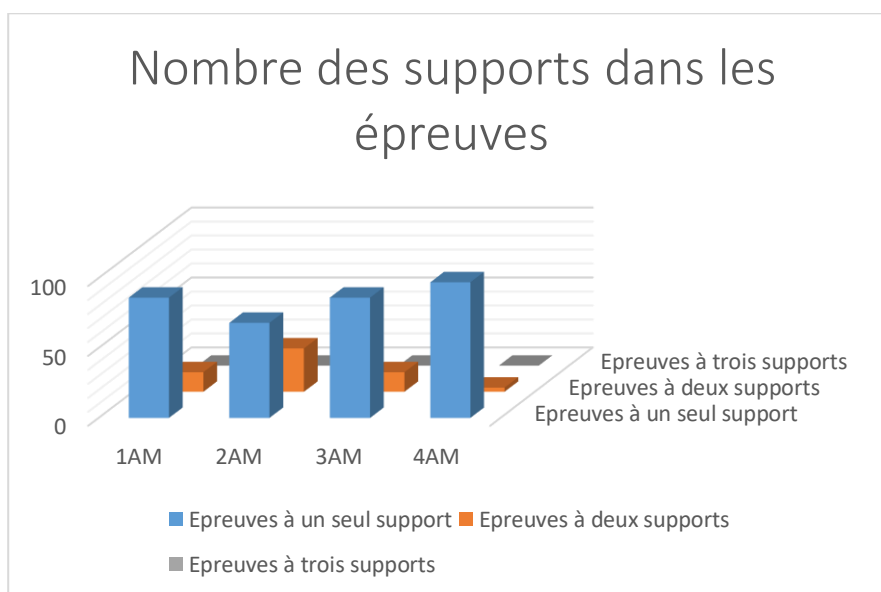
Les activités évaluant la compréhension et les ressources consistent en des questions de types variés sur les supports proposés. Mentionnons que ces deux composantes figurent dans une seule partie : compréhension de l'écrit. Quant à la seconde partie des évaluations

sommatives, elle consiste en des activités évaluant la compétence expressive des apprenants à travers des situations. Cette deuxième partie est intitulée *production écrite, expression écrite, écriture* ou *situation d'intégration*. De toutes ces appellations, celle de situation d'intégration nous interpelle car elle évoque la même confusion terminologique antérieurement abordée, au niveau des instructions officielles, étant donné que les situations évaluatives sont distinctes des situations d'intégration par leurs objectifs. Loin de cette complexité terminologique, il semble que les outils évaluatifs introduits en évaluation du FLE au cycle moyen varient entre ceux qui relèvent de la nouvelle approche (les situations) et ceux qui ne sont plus préconisés actuellement (les items). Cependant, il est important de vérifier à quel degré les situations conçues par les enseignants correspondent aux situations préconisées actuellement.

Les résultats exposés ci-dessus révèlent que les évaluations sommatives adoptent une instruction essentielle préconisée par les documents officiels : le recours aux situations en évaluation du FLE. Cet élément qui révèle une certaine congruence entre ce qui est réellement fait et ce qui doit être fait -et donc la validité du corpus étudié- ne dissimule pas la non-conformité à deux indicateurs de validité : l'évaluation séparée des compétences et celle des ressources. Étant donné que cette évaluation requiert l'introduction de supports variés ; nous en étudierons la validité dans les éléments suivants.

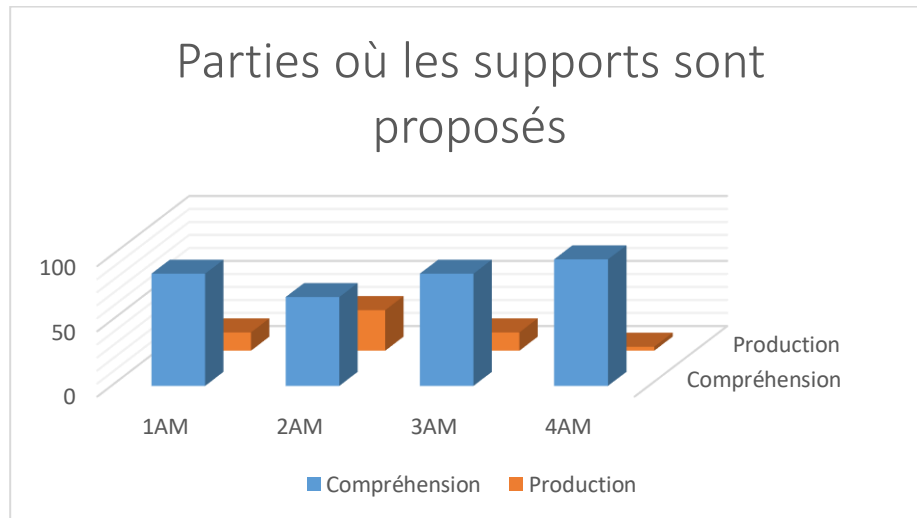
## 2.2. Les supports : quelle conformité aux instructions ?

Le premier indicateur dont nous présentons les résultats est le nombre de supports que les épreuves analysées intègrent. Pour ce faire, le graphique suivant représente les variables retenues : épreuves à support unique, épreuves à deux supports et épreuves à plus de deux supports.



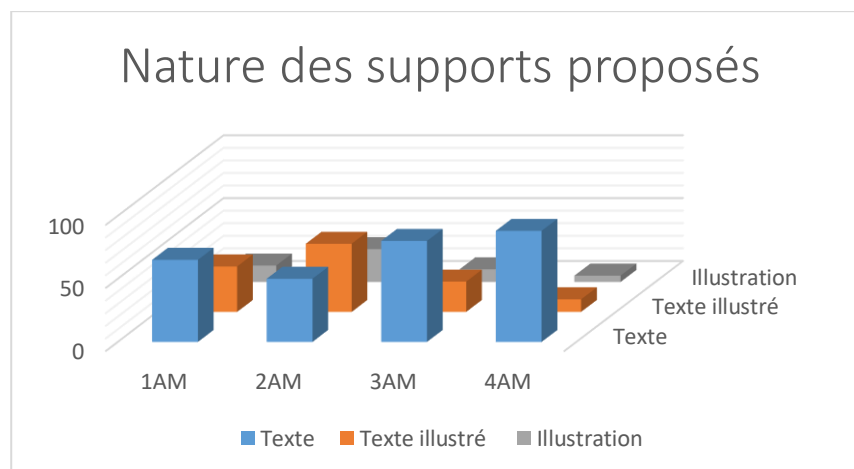
Graphique 9: Nombre des supports dans les épreuves des différents niveaux

Du graphique ci-dessus, il ressort que la quasi-totalité des épreuves analysées ne contiennent qu'un seul support alors qu'un nombre ostensiblement inférieur contient deux supports. Néanmoins, aucune épreuve ne comporte plus de deux supports. Il est important de préciser dans quelle partie ces supports sont introduits et d'en déterminer la nature pour comprendre pour quelles raisons les épreuves de 2AM notamment ; ainsi que ceux de 1AM contiennent deux supports. Vérifions d'abord dans quelle partie ceux-ci sont introduits à travers le graphique suivant :



**Graphique 10: Les parties des épreuves qui contiennent des supports**

Ce graphique semble identique au précédent vu le lien étroit entre les deux. Ce sont, en effet, les épreuves contenant deux supports qui proposent aux apprenants des supports dans la partie évaluant la production. Les autres épreuves se contentent d'un seul support proposé dans la partie qui évalue la compréhension. Mentionnons que ce sont des supports variés tel que le graphique ci-dessous le montre.

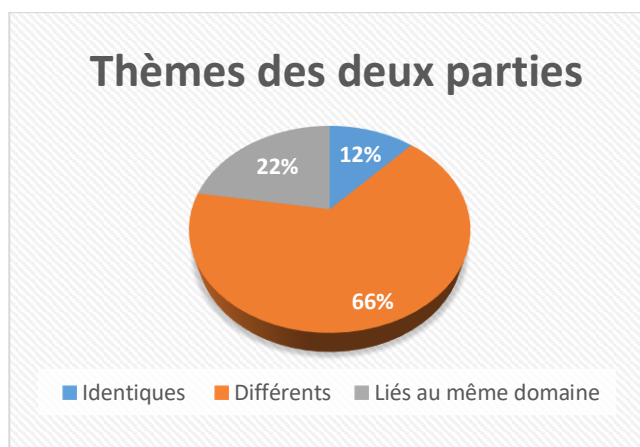


**Graphique 11: Nature des supports proposés dans notre corpus**

Trois types de supports existent dans les épreuves analysées : textes, textes illustrés et illustrations. Les textes sont très fréquents dans les épreuves de 4AM ainsi que celles de 3AM,

où le niveau des apprenants, plus avancés que les autres collégiens, permet le recours à des écrits non-illustrés. S'agissant des deux premiers niveaux du cycle (1AMet2AM), nous notons une présence importante des textes illustrés et des illustrations. Les enseignants les jugent plus adaptés au niveau des apprenants étant donné que les supports iconiques leur offrent des précisions supplémentaires. Il est évident que cet apport du support iconique ne retient pas notre attention à ce niveau : ce sont les textes illustrés et les illustrations, plus abondants dans les épreuves de 2AM, qui nous interpellent. Contrairement à ce que nous avons supposé, ces deux supports sont plus présents dans les épreuves de 2AM que celles de 1AM ; ce qui s'explique par la complexité du type textuel étudié : le texte narratif. Estimant difficile la lecture des textes qui relèvent de ce type, les enseignants introduisent des illustrations représentant souvent les protagonistes du récit en vue d'aider leurs apprenants à accéder au sens. Pour les illustrations sélectionnées en 1AM ; elles renvoient généralement aux thèmes abordés.

Nous devons mentionner que les supports proposés dans la deuxième partie des épreuves consistent en des illustrations représentant les objets abordés, les personnalités présentés, les personnages de la narration... Aucun texte ou texte illustré n'est recensé. Étant donné que les situations évaluant la compétence expressive ne contiennent pas de supports textuels sont la charge informative est indéniablement plus considérable, vérifions si les supports de la première partie pourraient être utiles à la seconde, en analysant le graphique qui suit :



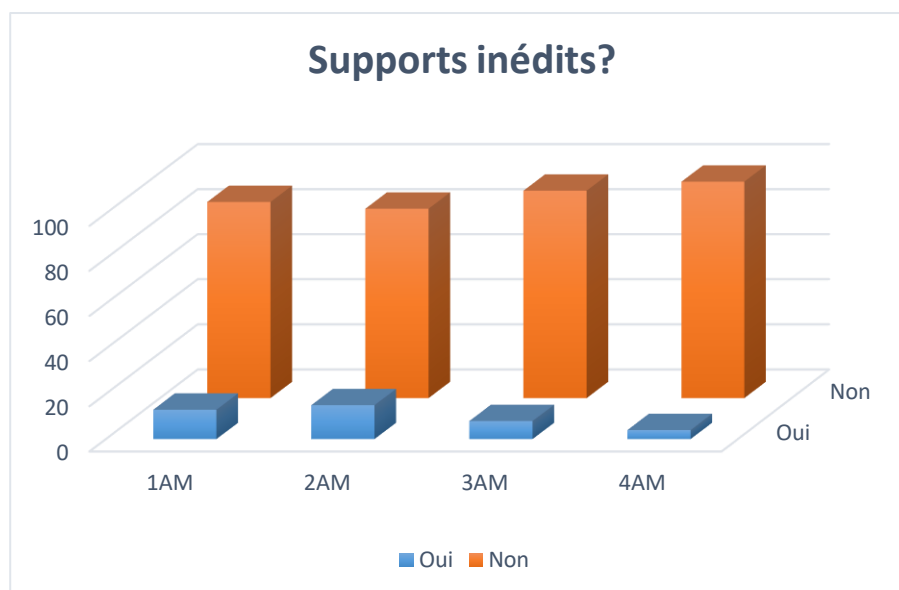
**Graphique 12: Thèmes des deux parties (Compréhension et Production)**

Nous avons préféré, à ce niveau de notre étude, aborder les thèmes traités dans les deux parties à travers trois variables précises. Comme le montre le graphique ci-dessus, le taux le plus important des thèmes des deux parties constituant les épreuves du FLE au cycle moyen revient aux thèmes différents. Autrement dit, le thème abordé dans la première partie et celui de la seconde sont si différents que les supports proposés en compréhension ne peuvent aucunement fournir aux apprenants les informations dont ils auront besoin. Le second taux,

nettement plus inférieur, représente les thèmes liés au même domaine. En dépit de ce lien, les éléments abordés dans la première partie sont souvent distincts de ceux de la seconde : les apprenants ne peuvent plus en extraire certaines informations utiles à leurs productions. Enfin, une minorité des thèmes sélectionnés sont identiques : la consigne de la production demande aux apprenants de traiter les mêmes éléments abordés dans le texte proposé. Dans ce cas-ci, les apprenants produisent des textes cohérents, répondant à la consigne, en reprenant partiellement ou intégralement les textes proposés en compréhension ; même si les enseignants les sanctionnent lors de la correction. Il semble ainsi que les enseignants sélectionnent intentionnellement des supports permettant une évaluation séparée des compétences ciblées, à travers des supports inutiles à la seconde partie constituant les épreuves analysées.

Les trois indicateurs ci-dessus révèlent que les supports, composantes essentielles des situations évaluatives, ne sont proposés que dans la première partie pour que les apprenants puissent répondre aux questionnaires de compréhension. Rares sont, en effet, les situations évaluatives qui comprennent des supports. Si ces supports existent, ils se réduisent à des illustrations représentant les thèmes à traiter (*des inventions technologiques en 1AM, des personnalités en 3AM, des fléaux sociaux en 4AM ...*), les personnages à décrire (*portraits en 2AM*)... ; ou déclenchant l'expression si celles-ci concrétisent quelques événements d'un récit (*illustrations proposées en 2AM pour que les apprenants racontent les fables du lièvre et la tortue, le corbeau et le renard...*). Bien que ces indicateurs révèlent que les situations évaluatives ne sont pas suffisamment conformes aux exigences et donc non valides pour une évaluation authentique des compétences des apprenants, nous poursuivons dans ce qui suit la description des supports que contient notre corpus. A travers cette description, nous montrerons si ces supports sont inédits, authentiques et d'une longueur adaptée au niveau des apprenants.

Pour déterminer si les supports sont inédits, nous vérifions s'ils figurent dans les manuels scolaires. Les supports inédits sont alors ceux qui ne sont pas déjà traités en classe de FLE, ou au moins qui n'existent pas dans les manuels scolaires. Il est important de mentionner que cette vérification ne portera pas exclusivement sur les manuels de deuxième génération étant donné que leur parution progressive a conduit les enseignants -soucieux de se conformer aux instructions officielles- d'adapter les anciens manuels aux exigences de la seconde réforme. En effet, les devoirs et les compositions qui relèvent de la deuxième génération sont en vigueur même avant la parution tardive de certains manuels (4AM). Le graphique ci-dessous, qui précise si les supports de ces devoirs et compositions figurent dans les manuels scolaires, présente les résultats des deux variables retenues (oui) et (non).

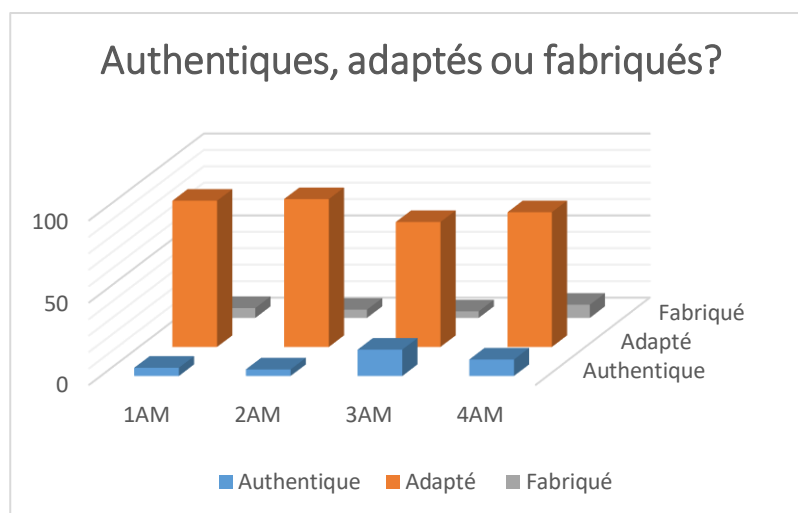


**Graphique 13: Les supports des épreuves : entre inédits et déjà exploités**

Ce graphique-ci montre que les enseignants recourent notamment à des supports inédits pour une évaluation correcte des compétences de leurs apprenants. Éviter les supports déjà exploités en classe semble évident étant donné que l'évaluation se transforme en une activité de restitution évaluant la capacité des apprenants à restituer des réponses à des questions antérieurement traitées. Bien que les enseignants puissent sélectionner des questions nouvelles, la proposition de textes déjà travaillés demeure inutile étant donné qu'une évaluation réelle devrait s'appuyer sur des textes nouveaux.

Le nombre de ces supports textuels nouveaux, que les enseignants exploitent en évaluation des compétences, est plus important dans les niveaux avancés (3AM et 4AM), ce qui traduit le souci des enseignants à obtenir des résultats plus révélateurs des compétences de leurs apprenants. Mais, ces supports inédits que les épreuves contiennent sont-ils authentiques ?

Les résultats collectés en réponse à cette question sont répartis en trois variables (authentiques, adaptés ou fabriqués) et présentés ci-dessus :



**Graphique 14: Les supports de notre corpus : authentiques, adaptés ou fabriqués ?**

Alors qu'une évaluation conforme aux exigences de l'APC préconise le recours à des documents authentiques en évaluation, les épreuves du cycle moyen continuent à privilégier les textes adaptés, leur accordant ainsi une certaine supériorité. C'est en 1AM et en 2AM que ces supports adaptés sont les plus fréquents étant donné que les enseignants procèdent à des modifications, parfois profondes, visant une adaptation de ces supports aux niveaux des apprenants évalués. Pour les autres niveaux, où les textes adaptés sont également plus fréquents, les évaluations analysées révèlent une présence plus importante des documents authentiques. Ceux-ci sont clairement plus nombreux en 3AM vu que la nature du contenu retenu (comme les faits divers) autorise les enseignants à sélectionner des supports répondant à certaines conditions : la complexité linguistique, la densité informative, la longueur... Le plus frappant est le nombre des documents fabriqués qui équivaut presque à celui des documents authentiques. La présence de ces documents fabriqués, déconseillés vu les incorrections multiples qu'ils pourraient contenir, s'explique par la volonté de proposer des supports accessibles<sup>112</sup> rassemblant des éléments précis à évaluer par les enseignants. Cette raison est valable pour les supports adaptés que les enseignants modifient pour orienter leurs évaluations vers des éléments précis. En effet, les enseignants ont tendance à proposer des textes dont la structure est identique à celle des supports des manuels scolaires pour que les apprenants puissent les comprendre en appliquant des démarches connues. Pour illustrer cette situation,

<sup>112</sup>Les enseignants -quand ils élaborent des supports pour évaluer les compétences de leurs apprenants en FLE- les écrivent dans une langue simple pour que ces derniers puissent les comprendre. Pour les textes adaptés, ils substituent les mots difficiles, inusités...par des mots simples et fréquents ; simplifient les structures grammaticales... Ces changements que les enseignants apportent ne sont pas toujours souhaitables étant donné qu'ils omettent la possibilité de découvrir de nouveaux mots et de nouvelles constructions et d'apprendre la langue dans toute son authenticité.

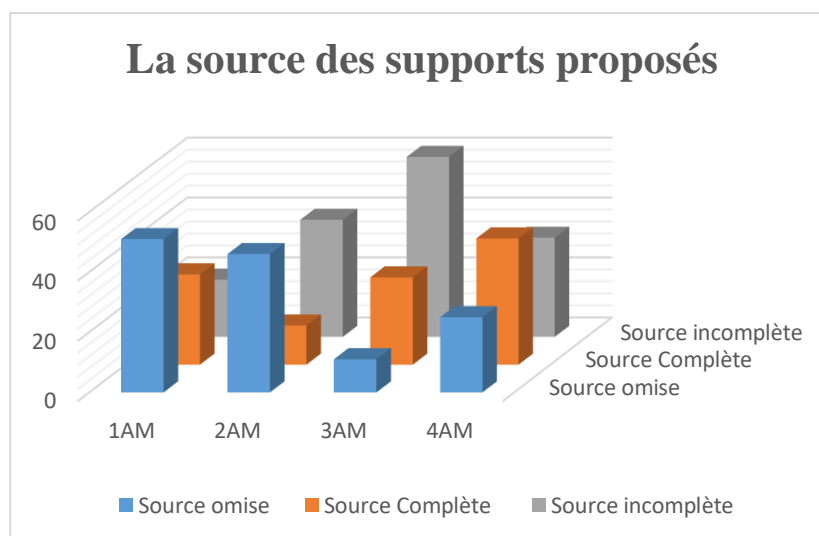


nous citons les textes proposés en 4AM et qui se caractérisent par une structure particulière marquée par la présence intentionnelle des connecteurs logiques introduisant les arguments. Ainsi, les enseignants extraient des textes variés et y insèrent des articulateurs marquant les arguments et les exemples, pour les rendre identiques à une structure particulière qui ne caractérise pas tous les textes relevant du type argumentatif. En effet, en opérant des modifications jugées nécessaires, les enseignants élargissent les écarts entre situations scolaires et situations authentiques étant donné que les supports relevant des divers types programmés ne se présentent pas dans la vie sociale sous une structure figée.

Cette structure figée obtenue suite à l'adaptation des textes, c'est-à-dire leur modification, conditionne les apprenants à privilégier une lecture sélective, orientée vers les outils linguistiques ; ce qui aboutit parfois à un repérage mécanique de certains éléments essentiels du texte. C'est le cas des apprenants de 4AM qui arrivent souvent à identifier les arguments et les exemples sans une réelle compréhension grâce aux connecteurs logiques qui les introduisent. Dans cette optique, le recours à ces textes adaptés, fidèles à des modèles précis, pourrait fausser les résultats des évaluations scolaires étant donné que les réponses correctes des apprenants à quelques questions de repérage ne refléteraient pas toujours leur compréhension de ce qui est lu. Les enseignants doivent alors proposer des textes authentiques, que nous opposons aux textes adaptés, en évaluation du FLE pour évaluer effectivement la compétence de compréhension chez leurs apprenants.

Il est important de préciser que les enseignants opèrent parfois des modifications si profondes qu'ils obtiennent des textes nouveaux qui résument les textes originaux ou en présentent quelques fragments enchaînés différemment. Une telle réalité suscite des interrogations quant aux modifications qui ne défigurent pas ces textes et qui pourraient donc être admises.

Pour indiquer que les textes soumis à leurs apprenants ne sont pas authentiques, les enseignants mentionnent « texte adapté », sans toujours citer la source ; comme le montre le graphique qui suit :



**Graphique 15: Les sources dans les textes proposés dans les épreuves**

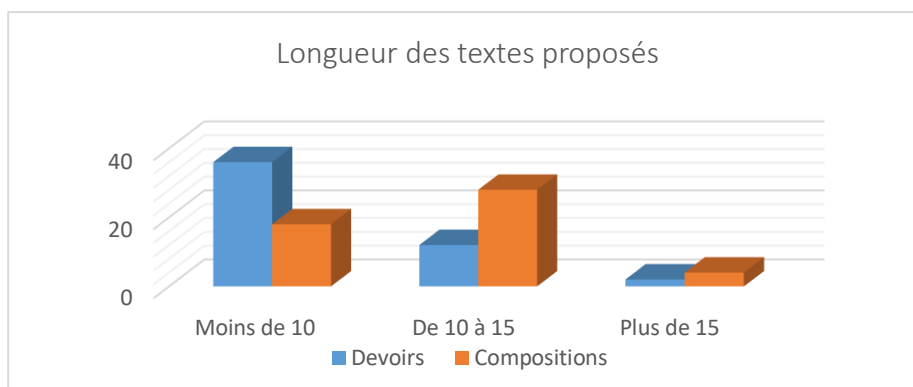
Le graphique révèle que les résultats varient d'un niveau à un autre. En effet, les deux premiers niveaux se caractérisent par le recours à des supports textuels dont les sources sont omises ; ce qui n'est pas le cas des niveaux plus avancés où les sources des textes sont mentionnées, complètes ou incomplètes. Étant donné que les enseignants se conforment aux prescriptions préconisant des textes authentiques, nous pouvons croire qu'ils élaborent et sélectionnent plus soigneusement les supports proposés avec les niveaux avancés. Cette remarque ne répond pas à la question suivante qui s'impose à ce niveau de notre étude : pourquoi les sources sont-elles omises ?

La première réponse que nous pouvons déduire de la représentation graphique ci-dessus est que ces textes sont extraits de références qui n'en citent pas la source (des pages web, des documents électroniques, des devoirs, des compositions...). Nous estimons logique cette justification car nous constatons que certains textes sont proposés dans plusieurs devoirs et compositions ; ce qui montre que les enseignants puisent leurs supports même dans les épreuves existantes. Si les enseignants ne citent pas toujours les sources, c'est parce que celles-ci ne sont pas mentionnées dans les références consultées. Cette omission ne pourrait pas être intentionnelle vu que les enseignants précisent les sources dont ils disposent.

Lorsque les sources sont citées, elles ne sont pas toujours complètes car les enseignants se contentent de mentionner l'auteur et l'ouvrage (intitulé), sans fournir des indications qui permettent aux apprenants de vérifier les informations fournies ou découvrir plus de renseignements sur les sujets traités. Nous constatons en effet que les supports proposés en 3AM sont souvent suivis de leurs sources, même incomplètes. Donc, les enseignants sont conscients de la nécessité de mentionner les sources des supports sélectionnés et leur omission n'est pas due à une négligence ; mais plutôt à leur absence dans les références consultées.

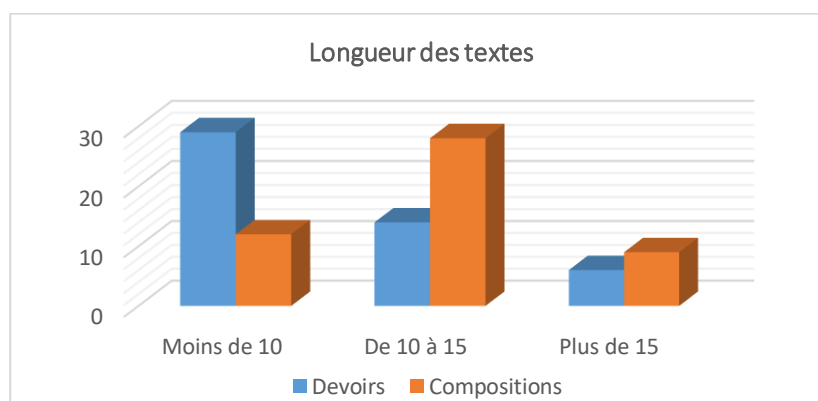
La dernière précision sur les supports textuels sélectionnés dans les épreuves collectées concerne leur longueur. Contrairement aux autres indicateurs où les résultats des devoirs et des compositions étaient présentés indissociablement, les graphiques suivants les dissocient vu les écarts constatés.

Examinons le premier graphique qui représente la longueur des textes de 1AM.



**Graphique 16: Longueur des supports textuels proposés en 1AM**

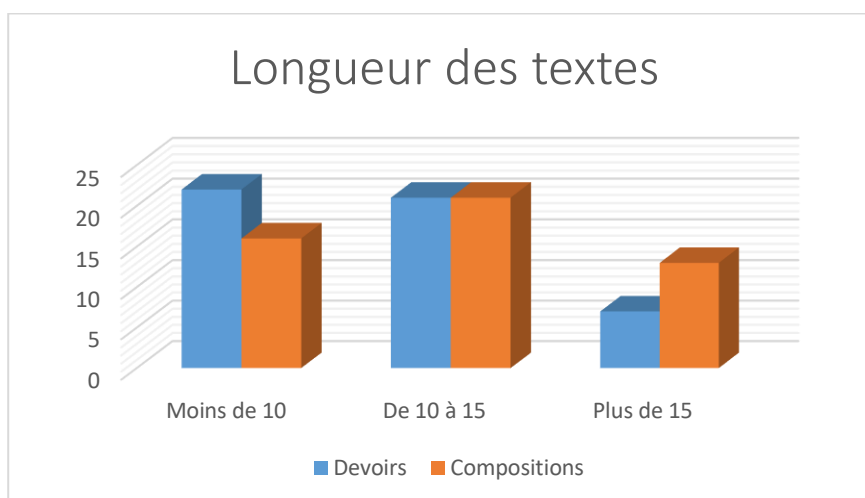
A travers le graphique, nous constatons que les textes dont la longueur est inférieure à 10 lignes sont fréquents dans les devoirs où les enseignants se permettent quelques transgressions aux exigences vu la durée impartie (aux devoirs notamment). Ce cas ne s'applique pas à la deuxième variable, qui renvoie à la longueur-norme, où le nombre des compositions est clairement plus important. Ainsi, les textes sélectionnés dans les compositions de 1AM sont d'une longueur conforme aux exigences et adaptés au niveau des apprenants ainsi qu'au temps impartie à l'épreuve. Ce constat sur les épreuves de 1AM est valable à celles de 2AM dans la mesure où :



**Graphique 17: Longueur des textes proposés en 2AM**

Les textes dont la longueur est inférieure à la norme sont fréquents dans les devoirs alors que ceux dont la longueur est conforme à la norme sont plus nombreux dans les compositions ; ce qui confirme que les enseignants ont tendance à obéir davantage aux normes lors de la conception des compositions.

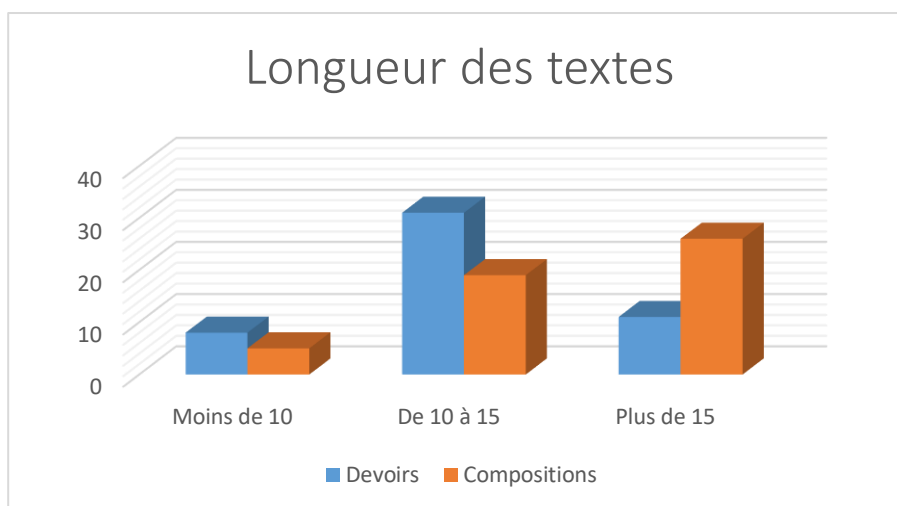
Ces affirmations ne s'appliquent pas cependant aux épreuves de 3AM et de 4AM. Ainsi, à travers ce graphique :



**Graphique 18: Longueur des textes proposés en 3AM**

Nous constatons que les devoirs, de 3AM, sont plus nombreux dans les deux premières variables. Peu d'enseignants proposent des supports dont la longueur dépasse 15 lignes en raison de la durée des devoirs qui ne permet pas aux apprenants de lire des textes longs. Pour ce qui est des compositions, les supports textuels intégrés sont souvent d'une longueur conforme à la norme. Néanmoins, certaines épreuves intègrent des textes dont la longueur ne dépasse pas 10 lignes ; ce qui reflète des choix purement personnels liés au niveau des apprenants et aux objectifs visés par les enseignants. Soulignons la présence de quelques textes longs notamment dans les compositions, étant donné que la durée des épreuves le permet.

Pour les épreuves de 4AM, la situation est plus claire étant donné que les textes sont fréquents dans les deux dernières variables. Ainsi, les devoirs intègrent souvent des textes dont la longueur correspond à la norme. Cependant, certaines épreuves contiennent des textes longs en dépit du temps qui leur est imparti. S'agissant des compositions, les textes longs sont plus fréquents car le niveau des apprenants et la durée le permettent. Nous notons néanmoins que certains enseignants se conforment aux exigences en sélectionnant des textes d'une longueur nécessaire. Ces affirmations s'appuient sur le graphique qui suit :



**Graphique 19: Longueur des textes proposés en 4AM**

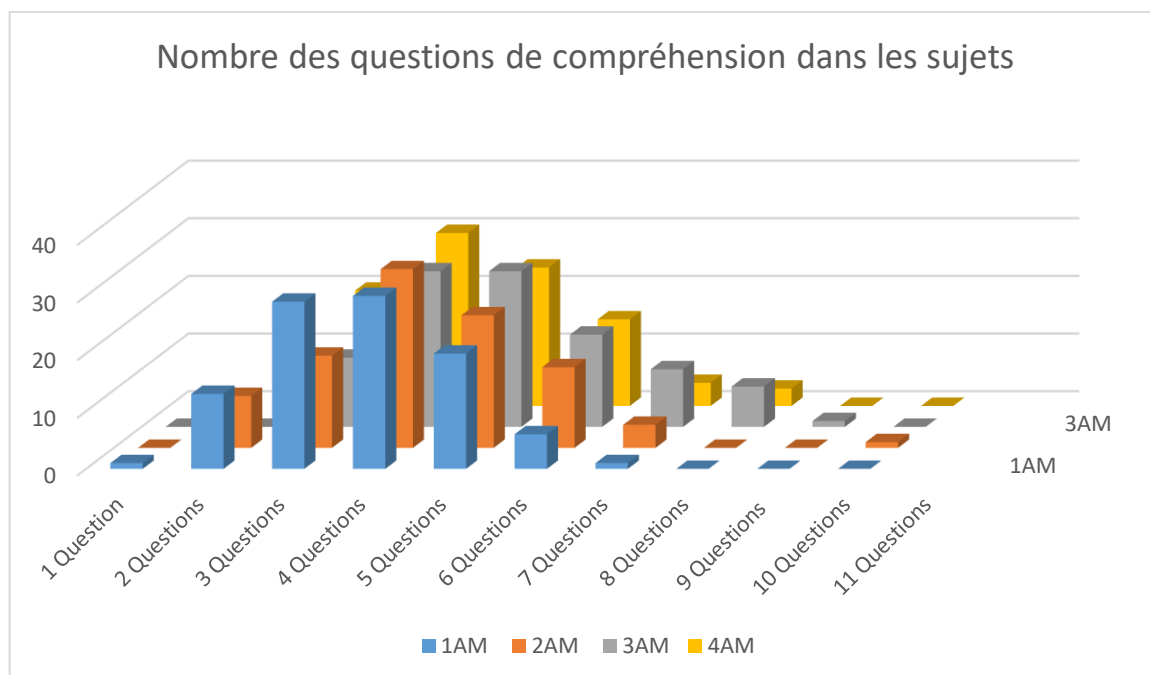
De ce qui précède, il semble que la longueur des textes inclus dans les épreuves analysées dépend des choix des enseignants qui, en fonction de leurs objectifs, optent pour une longueur particulière. Il est clair, cependant, que leurs choix tiennent compte de deux contraintes essentielles. La première est le niveau des apprenants qui détermine la longueur des textes à proposer : les textes longs sont plus nombreux dans les niveaux avancés du cycle. La deuxième contrainte est temporelle : la durée impartie aux épreuves détermine la longueur des textes sélectionnés. C'est pour cette raison que des textes plus longs sont retenus dans les compositions, contrairement aux devoirs où les enseignants optent pour des supports moins longs. Ces deux facteurs ne fournissent pas cependant une explication assouvissant nos interrogations étant donné que certains résultats obtenus ne justifient plus nos affirmations, comme c'est le cas de la 3AM où les supports des devoirs et les compositions, en dépit de leurs durées distinctes, sont d'une longueur similaire. Une telle situation suppose une sélection parfois irréfléchie des supports et une évaluation privilégiant les outils de mesure.

En somme, il est clair que les épreuves évaluatives constituées de deux parties intègrent souvent un support unique dans la partie évaluant la compréhension, ce qui rend les situations évaluatives lacunaires. En effet, celles-ci ne comportent pas de supports informatifs, qui en constituent une composante essentielle. D'ailleurs, les supports de la première partie, qui ne répondent pas à toutes les exigences, ne sont plus utiles à la seconde vu que les thèmes évoqués sont souvent distincts. Bien que ce manque de supports, dans les situations évaluatives, en altère la validité, nous ne pouvons les juger sans une analyse minutieuse des outils de mesure et de jugement.

### 2.3. Les outils de mesure

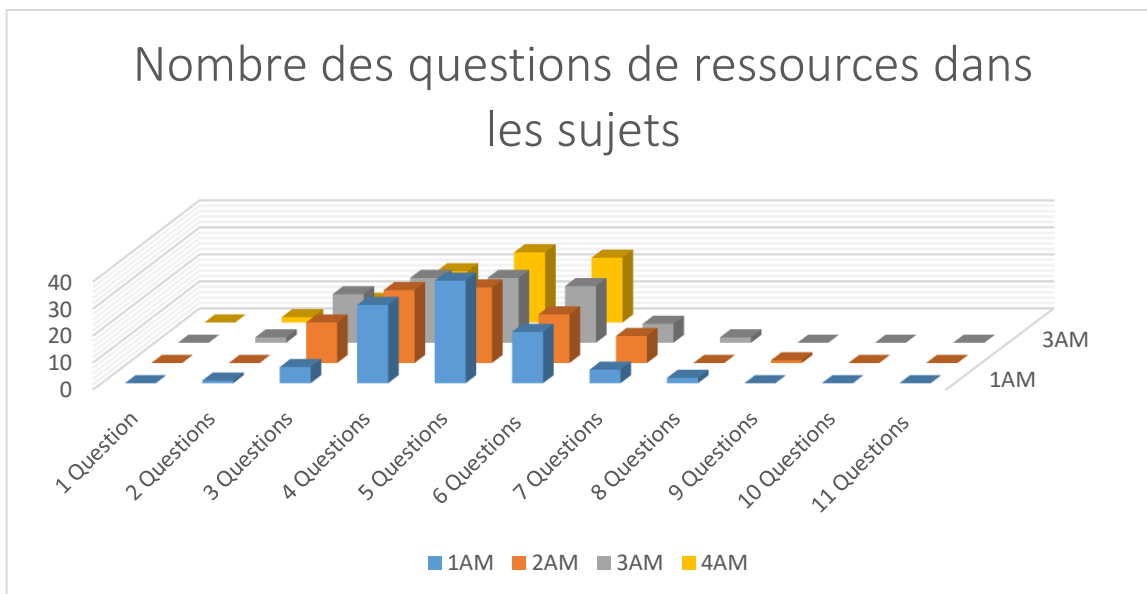
Les résultats liés aux instruments de mesure ne sont pas présentés suivant leur ordre dans la grille d'analyse introduite dans notre étude. Certains seront présentés en parallèle pour des comparaisons que nous jugeons pertinentes dans notre étude. C'est le cas du nombre des questions évaluant la compréhension et de celui des questions qui évaluent les ressources.

Examinons le graphique suivant :



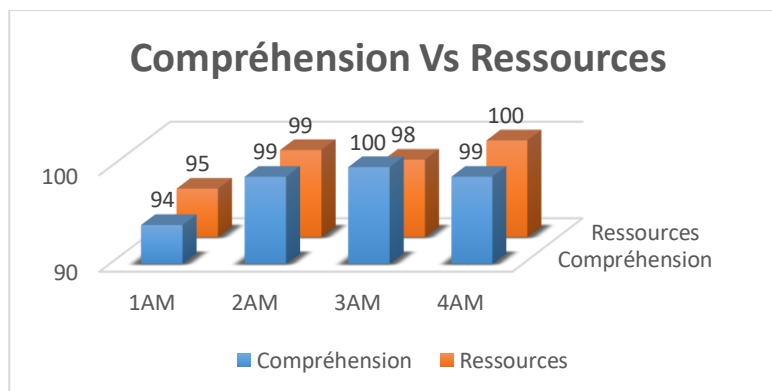
**Graphique 20: Nombre de questions évaluant la compréhension dans les sujets analysés**

Dans ce graphique, nous présentons les sujets analysés selon leur nombre de questions proposées pour évaluer la compréhension. Il est clair que les enseignants optent fréquemment pour un nombre qui varie de trois (03) à six (06) questions pour évaluer la compréhension des supports écrits. Il n'existe pas de différences significatives entre devoirs et compositions même si le temps imparti aux deux types n'est pas identique. Pour le nombre des questions qui évaluent les ressources, nous le présentons dans le graphique suivant :



**Graphique 21: Nombre des questions évaluant les ressources dans les sujets analysés**

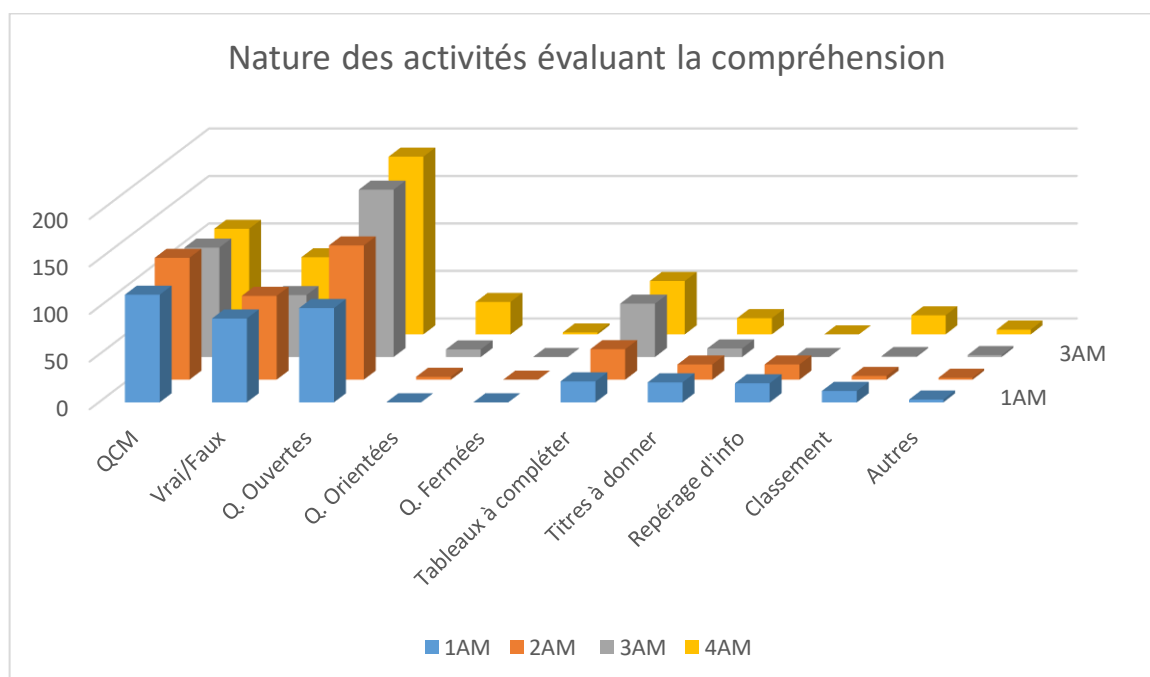
Contrairement à nos attentes, ce graphique qui représente le nombre des questions qui évaluent les ressources paraît conforme au précédent étant donné que le nombre fréquent varie de trois (03) à six (06) questions. Cependant, il semble que le nombre des questions évaluant les ressources équivaut celui des questions qui évaluent la compréhension, un constat que nous vérifions à travers le graphique ci-dessous qui confronte les deux catégories de questions.



**Graphique 22: Comparaison du nombre des questions des deux catégories (Compréhension et Ressources)**

Ce graphique-ci confirme notre constat et montre que le nombre des questions qui évaluent la compréhension est quasi-identique à celui des questions portant sur les ressources. Un tel résultat remet en question la validité des épreuves conçues pour évaluer les compétences des apprenants en FLE qui ne doivent pas accorder tant de valeur aux questions évaluant les ressources. Cette question de validité ne se clôt pas cependant par le volume des ressources dans les sujets analysés : la nature des activités est à examiner pour déterminer si elles évaluent réellement la compréhension et les ressources.

Le graphique qui suit présente les principales activités que les enseignants du cycle moyen en Algérie introduisent pour évaluer la compréhension des supports proposés.



**Graphique 23: Représentation des types d'activités évaluant la compréhension**

Cette représentation graphique révèle que trois activités sont fréquentes dans les épreuves évaluant la compréhension des collégiens en FLE : les questions ouvertes, les questions vrai/faux et les questions à choix multiples (QCM). Ces activités sont-elles valides pour évaluer la compréhension ?

Les questions ouvertes -qui mobilisent les opérations cognitives de haut niveau- évaluent certainement la compréhension. Cependant, deux facteurs en biaisent la validité. Le premier est que les questions formulées par les enseignants pourraient être ambiguës, imprécises ou incorrectes, induisant par-là les apprenants en erreur : une réponse incorrecte ne résulte pas toujours d'une mauvaise compréhension du texte lu. C'est pour cette raison que la conception des examens est une tâche délicate qui requiert la coordination des enseignants-concepteurs. Le second facteur est lié à la compétence expressive que ces questions ouvertes mobilisent. Les apprenants pourraient formuler incorrectement leurs réponses en raison de leur compétence restreinte en FLE ; ce qui ne révèle pas la compréhension réelle que les apprenants ont construite des textes lus. Notons une évolution progressive du nombre de ces questions dans les niveaux du cycle liée à une évolution présumée des compétences linguistiques des apprenants.

Contrairement aux questions ouvertes, les questions présentant des alternatives (QCM, Vrai/Faux...) ne requièrent pas une expression des réponses : elles évaluent réellement la



compréhension des textes. D'autres avantages leur sont également attribués. Leclercq (1986) affirme que les QCM assurent une évaluation globale des supports textuels sélectionnés et une correction objective des réponses sélectionnées par les apprenants. Ces avantages ne peuvent néanmoins exclure des facteurs qui biaisent les réponses comme la tricherie et le hasard étant donné que sélectionner les réponses nécessaires ne traduit pas toujours les compétences des évalués. A ces facteurs, nous jugeons nécessaire d'inclure la formulation de la question par les enseignants qui doivent sélectionner des distracteurs présentant une réelle confusion. Cependant, cette exigence fait défaut dans les QCM que les épreuves du cycle moyen contiennent. En voici quelques exemples :

*Le thème de ce texte est :*  
*Le racisme, Le travail humanitaire, La tolérance.*  
*Le problème traité dans ce texte est :*  
*La solidarité, La violence, La discrimination.*  
*Dans ce texte, on parle :*  
*D'un moyen de transport, De la pollution, D'une invention.*  
*La reine habitait dans :*  
*Un château, Une maison, Une cabane.*

Ces distracteurs illustrent nos propos et révèlent que ces solutions ne présentent pas de confusion notamment que le questionnaire porte sur un support unique, ce qui ne serait pas identique si plusieurs supports sont sélectionnés. En dépit de cette critique que nous formulons, nous ne pouvons nier que ces solutions respectent les règles qui régissent la rédaction des QCM que Leclercq explicite dans son précieux ouvrage *La conception des Questions à Choix Multiples*. Entre autres, citons : l'indépendance sémantique, le degré de complexité...

Quant aux questions Vrai/Faux, elles présentent les mêmes avantages que les QCM et elles sont biaisées par les mêmes facteurs. Pour les formuler, les enseignants extraient qui reprennent littéralement des éléments du texte ou introduisent des synonymes ou des antonymes, des adverbes de négation... Vu une telle formulation, les réponses correctes des apprenants deviennent peu indicatrices du degré de la compréhension des supports proposés. Voici des phrases –choisies au hasard- extraites des épreuves analysées :

*Les deux voyageurs aperçurent un ours.*  
*L'avion a été abaissé mercredi matin.*  
*Le stade est devenu un endroit dangereux.*  
*Le plastique est difficile à fabriquer.*

Ces exemples correspondent respectivement aux phrases suivantes dans les supports proposés (*deux voyageurs, faisant chemin ensemble, aperçurent un ours ; un avion militaire s'est écrasé, au sud d'Alger, mercredi matin ; le stade est devenu un champ de bataille dangereux ; il est facile à fabriquer*), ce qui illustre les cas précédemment cités. Il existe une

variété de cette question Vrai/Faux, rare dans notre corpus, où il est demandé aux apprenants de corriger les phrases incorrectes. En dépit de son utilité en évaluation de la compréhension, elle implique la compétence expressive créant ainsi des problèmes de validité.

A ces trois types fréquents dans notre corpus, s'ajoutent des activités comme les tableaux à compléter, les titres à donner... Dans les tableaux à remplir, des questions ouvertes et interdépendantes sont souvent proposées et les apprenants y répondent à partir des supports lus. Ces tableaux -quasi typiques des types textuels étudiés- sont parfois utiles pour évaluer la compréhension. Examinons le tableau ci-dessous qui évalue la compréhension des articles journalistiques :

<i>Quoi ?</i>	<i>Où ?</i>	<i>Quand ?</i>	<i>Conséquences</i>

Cependant, le tableau suivant qui évalue la compréhension des textes narratifs repère des informations peu indicatrices du degré de leur compréhension.

<i>Où ?</i>	<i>Quand ?</i>	<i>Personnage 1</i>	<i>Personnage 2</i>

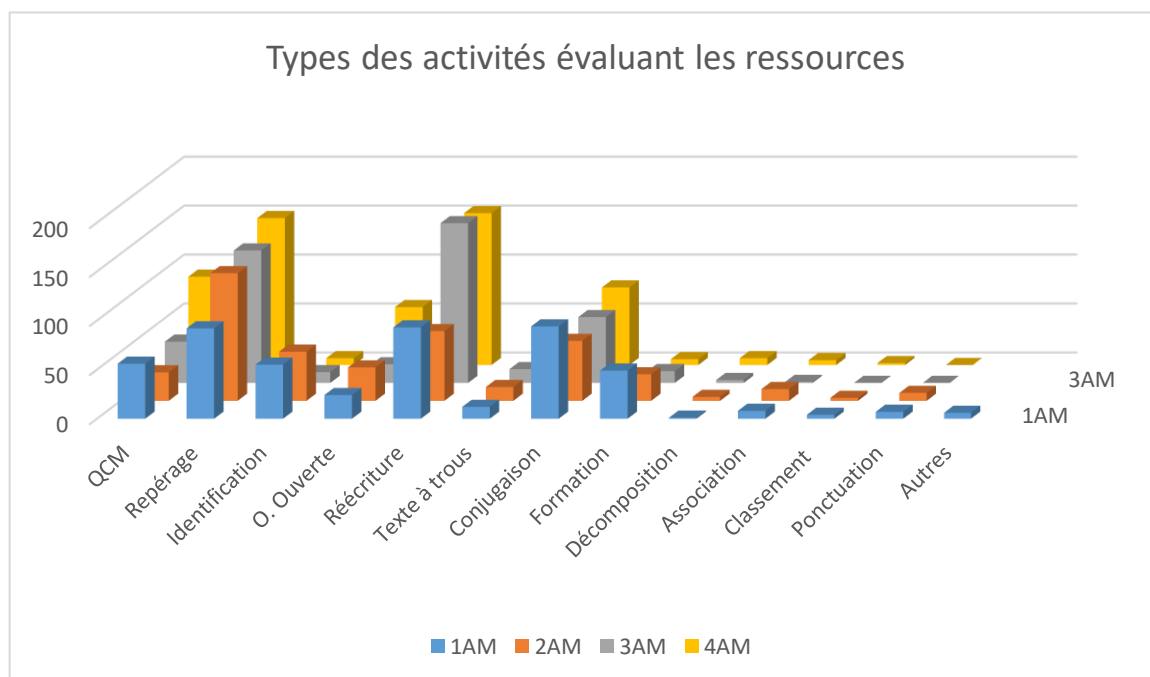
Bien que ces tableaux soient constitués de questions ouvertes, il est exigé des apprenants de repérer des éléments précis des supports proposés, tout comme dans les consignes de repérage. Le fait que ces questions ne requièrent que le repérage ou la formulation de points précis les rend valides pour évaluer la compréhension des apprenants. En dépit de leur validité, ces questions demeurent peu nombreuses dans les épreuves écrites analysées tout comme les questions orientées, les questions de classement, d'association, de remise en ordre, les textes à trous...

Pour les questions orientées (guidées) où les examinateurs demandent aux apprenants de justifier des informations fournies, elles sont plus nombreuses au niveau le plus avancé du cycle moyen (autrement-dit la 4AM) étant donné que le niveau des apprenants autorise l'introduction de ce type de questions. La justification de ces informations ne mobilise pas cependant la compétence expressive des apprenants, d'où leur validité, car les apprenants doivent les justifier par des phrases relevées des textes lus. Souvent, la consigne est ainsi formulée : *relève du texte une phrase qui montre que (...)* ; *justifie cette affirmation par une phrase du texte...* Il en est de même pour les questions de classement dont la validité est peu contestée vu que les éléments à classer sont listés et la tâche des apprenants consiste à les répartir dans les colonnes appropriées. Dans la deuxième variante de cette activité, les éléments à classer doivent être repérés par les apprenants eux-mêmes dans les supports sélectionnés, ce

qui n'en altère pas la validité ; qui serait affectée si les éléments à classer ne figurent pas dans ces supports.

Les activités d'association et de remise en ordre sont aussi valides que celles du classement vu que seule la compréhension est évaluée. Cependant, certains écarts pourraient en altérer la validité comme l'absence des propositions dans les textes à trous constatée dans certaines épreuves.

Bien que certains types évoqués jusqu'ici semblent valides car ils évaluent la compréhension sans intégrer la variable expressive, les épreuves analysées ne pourraient être déclarées valides sans prendre en considération tous les indicateurs de validité que notre grille analytique comprend. Ces types sont-ils les mêmes que les enseignants introduisent en évaluation des ressources ? Cette question, nous y répondons à travers le graphique qui suit :



**Graphique 24: Répartition des activités évaluant les ressources**

En dépit de leur variété, certains types semblent plus courants que d'autres en évaluation des ressources. Ainsi, les activités de repérage et celles de réécriture (notamment en 3AM et en 4AM) sont les plus introduites par les enseignants pour évaluer la maîtrise des éléments grammaticaux enseignés. Quelles composantes ces activités évaluent-elles ?

Les activités de repérage requièrent l'identification d'éléments précis dans le support sélectionné : *synonymes, antonymes, mots de même famille, mots du même champ lexical...*, et

permettent notamment une évaluation des contenus lexicaux enseignés. Elles sont étroitement liées aux activités d'identification, représentées séparément dans le graphique ci-dessous, où les enseignants demandent aux apprenants d'identifier les mêmes éléments. Contrairement aux premières, les secondes ne nécessitent pas la recherche dans un texte car les éléments sont clairement indiqués dans le texte ou une phrase précédant la consigne. Estimées plus complexes, les activités de repérage sont de plus en plus nombreuses dans le cycle moyen contrairement aux activités d'identification qui diminuent dans les niveaux les plus avancés. Cependant, ces activités dites complexes sont-elles si importantes pour les introduire en évaluation sommative ? Il est important de préciser que les activités d'identification évaluent également la syntaxe (*identifier le discours d'une phrase*), la conjugaison (*identifier les rapports logiques, les temps verbaux...*) et même l'orthographe (*identifier des homophones lexicaux...*).

Les activités de réécriture ne sont pas introduites pour évaluer une seule composante grammaticale : elles évaluent la syntaxe, la conjugaison et même l'orthographe. Pour évaluer la syntaxe, les enseignants demandent aux apprenants de transformer des phrases indépendantes d'une voix à une autre (*de la voix active à la voix passive ou inversement*), d'un style à un autre (*du style direct au style indirect ou inversement*)... Ce qui est enseigné est jugé maîtrisé si les élèves réussissent à appliquer les règles apprises. En évaluation de la conjugaison, les élèves doivent réécrire des phrases dans des temps ou des modes différents (*réécris la phrase au futur simple*). La conjugaison est également évaluée à travers les activités de réécriture qui de prime abord visent une évaluation orthographique. *En réécrivant des phrases au pluriel, au singulier, au féminin ou au masculin*, les apprenants doivent conjuguer correctement les verbes à modifier.

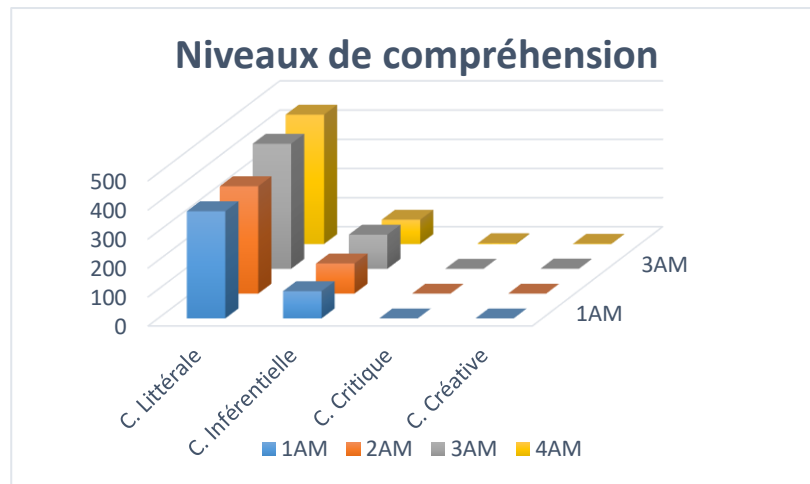
Cependant, ces activités de repérage et de réécriture sont-elles valides ? Évaluent-elles ce que les enseignants veulent évaluer ? Si nous ne tenons pas compte des exigences de l'approche actuellement en vigueur, ces activités deviennent appropriées car elles évaluent réellement ce que les enseignants veulent évaluer. Cependant, dans le cadre de l'APC, une telle évaluation est non valide car les savoirs et les savoir-faire des apprenants ne leur suffisent pas pour communiquer et agir en FLE. Il en est de même pour les autres activités que les épreuves analysées introduisent pour évaluer les diverses composantes. Ainsi :

- Les QCM : évaluent notamment le vocabulaire (*reconnaître un synonyme, un antonyme... dans une liste de mots*) et la syntaxe (*identifier le rapport logique et les mots qui l'expriment, identifier la nature des subordonnées...*) ;

- Les questions ouvertes : sont introduites surtout pour évaluer la syntaxe (*identifier les rapports logiques...*) et la conjugaison (*reconnaître les temps verbaux...*) ;
- Les textes à trous : sont très fréquents pour évaluer l'orthographe (*compléter par les homophones proposés...*) ;
- Les activités de formation des mots : évaluent le vocabulaire (*former des noms à partir de verbes, former des adverbes à partir d'adjectifs...*) ;
- Celles de décomposition : permettent d'évaluer le vocabulaire (*décomposer des mots dérivés*) et la syntaxe (*décomposer les phrases complexes en propositions*) ;
- Les activités de classement : évaluent le vocabulaire (*distinguer le vocabulaire mélioratif du péjoratif*) et la syntaxe (*distinguer les expansions des noms*)...

Cette évaluation –peu conforme aux exigences de l'APC - pourrait-elle nous informer sur les pratiques enseignantes ? Serait-elle la traduction de ce qui est enseigné et visé par les enseignants dans leurs classes de FLE ? Pour des réponses satisfaisantes à ces questions, nous exposons les résultats de deux indicateurs liés aux deux derniers : les niveaux de compréhension visés à travers les questions de compréhension et les niveaux cognitifs des items évaluant les ressources.

Pour les premiers, il semble d'après le graphique ci-dessous que les épreuves écrites que les enseignants introduisent pour évaluer les compétences en FLE privilégient la *compréhension littérale* des supports textuels proposés, d'où les réponses que les apprenants formulent sont souvent extraites littéralement. Il est clair que cette compréhension littérale est dominante même aux niveaux les plus avancés du cycle (4AM) qui devraient intégrer les autres niveaux de compréhension. Bien que les questions de ce type soient plus certaines lors de la correction, elles ne peuvent garantir une compréhension réelle des supports lus. Dans le cas où les enseignants ne requièrent pas une formulation correcte des réponses, les apprenants pourraient répondre correctement en réécrivant ou sélectionnant des phrases présentant des similitudes avec les énoncés des questions. Cependant, si les enseignants sanctionnent la correction linguistique des réponses formulées, la compétence expressive sera intégrée ; nuisant par-là à la validité de leur correction. Ces deux contraintes évoquées sont notamment liées aux questions ouvertes et aux QCM, plus fréquentes dans les épreuves analysées.



**Graphique 25: Les niveaux de compréhension évalués dans les épreuves analysées**

Pour les questions évaluant la compréhension inférentielle, elles portent souvent sur le genre, le type et la visée des supports textuels retenus et les apprenants, pour y répondre, ont à sélectionner ce qui est correct parmi les suggestions fournies (QCM). Outre les biais liés aux QCM, les distracteurs sélectionnés ne relèvent pas de ce qui est enseigné. Observons les distracteurs de la question suivante destinée aux apprenants de 1AM :

*Ce texte est de type :*

- *Narratif,*
- *Descriptif,*
- *Explicatif.*

Ces trois distracteurs représentent des types textuels enseignés respectivement dans les trois niveaux du cycle : 2AM, 3AM et 1AM. Ainsi, les suggestions ne présentent pas aux apprenants un contenu enseigné et intègrent des contenus non-enseignés ; ce qui ne suscite aucune confusion étant donné que les apprenants sélectionnent le type approprié sans une lecture profonde des supports proposés. Dans ce cas-ci, les biais résultent notamment de la formulation des questions par les enseignants qui doivent sélectionner des solutions plus appropriées.

Les derniers niveaux de la compréhension qui pourraient être évalués (la *compréhension critique* et *créative*) -qui ne sont pas exempts de biais- sont rares dans les épreuves analysées. En effet, nous ne recensons que quatre questions présentées dans les exemples suivants :

- *Cette description est-elle subjective ou objective ? Justifie ta réponse (Compréhension critique).*
  - *La conclusion est-elle optimiste ? (Compréhension critique).*
- *Donne ton point de vue sur le thème du texte (Compréhension créative).*

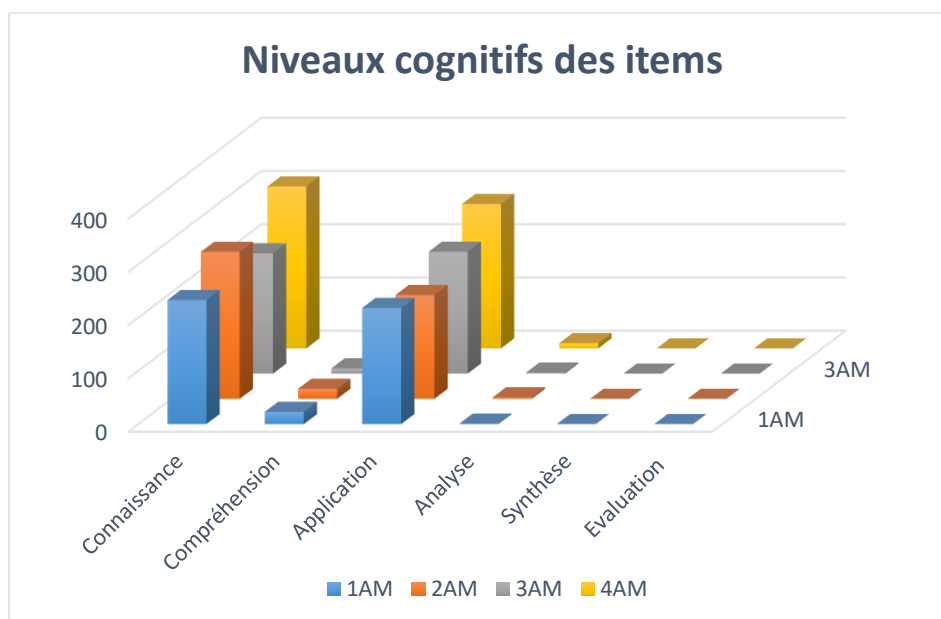
Cette analyse des niveaux de compréhension évalués dans les épreuves du cycle moyen révèle que les niveaux les plus hauts sont pratiquement négligés en évaluation du FLE. Ceci

signifie leur négligence dans les pratiques enseignantes qui semblent privilégier la compréhension littérale notamment que les enseignants évaluent ce qui est enseigné. Ces résultats liés aux niveaux de compréhension ne sont pas différents de ceux obtenus suite à la représentation graphique, ci-dessous insérée, des niveaux cognitifs des items évaluant les ressources. Cette représentation-ci traduit une rareté, voire une absence, des niveaux cognitifs supérieurs ; suscitant de nombreuses interrogations : pour quelles raisons les niveaux supérieurs sont-ils écartés ? Une telle évaluation est-elle conforme aux exigences ?

Contrairement aux niveaux supérieurs, les niveaux taxonomiques inférieurs sont les plus évalués au cycle moyen notamment la *connaissance* et *l'application*. Les privilégier en évaluation indique les croyances des enseignants-examineurs et trahit leurs pratiques enseignantes. Ceux-ci estiment maîtrisés les éléments grammaticaux enseignés si les apprenants les identifient dans des supports ou des phrases décontextualisées ; et s'ils les appliquent correctement dans de nouveaux contextes, ce qui explique la centration sur ces deux objectifs en enseignement et en évaluation. Voici des exemples extraits de notre corpus :

1. *Souligne le GN (Groupe Nominal) et encadre l'adjectif qualificatif :*
  - *Eviter les boissons sucrées.*
  - *Il faut prendre un repas varié.*
2. *Relève du texte une proposition subordonnée relative.*
3. *« La violence est une bête noire », confirme-t-il.*  
*Transforme cette phrase au style indirect.*

Ces consignes récurrentes dans les épreuves écrites évaluant les compétences des collégiens algériens en FLE révèlent une évaluation décontextualisée des ressources apprises et une valorisation de la maîtrise du métalangage grammatical, qui ne pourraient pas prédire les compétences réelles des évalués. Cette manière dont les ressources sont évaluées peu conforme aux exigences de l'APC nécessite une révision des pratiques enseignantes et évaluatives.



**Graphique 26: Les niveaux cognitifs des items évaluant les ressources**

Bien que cette évaluation soit peu conforme aux exigences, nous ne pouvons nier que les questions sélectionnées sont souvent représentatives des contenus enseignés dans la mesure où les objectifs visés en enseignement sont intégrés en évaluation. Vocabulaire, syntaxe, conjugaison et orthographe sont, en effet, évalués suivant leur ordre les séquences pédagogiques à travers des items dont le nombre est souvent aléatoire.

L'analyse des niveaux cognitifs évalués a indirectement anticipé les résultats d'un autre indicateur que nous analyserons dans notre étude sur la validité des épreuves du FLE au cycle moyen : le lien entre les questions et les supports sélectionnés. Les questions posées par les évaluateurs sont-elles alors liées aux supports proposés ? La réponse à cette interrogation est affirmative pour les questions évaluant la compréhension et négative pour celles portant sur les ressources. Si les questions évaluant les divers niveaux de compréhension sont en lien étroit avec les supports sélectionnés, celles qui portent sur les ressources sont souvent décontextualisées<sup>113</sup>. Pour évaluer la syntaxe, la conjugaison et l'orthographe, les enseignants proposent des questions entièrement indépendantes des textes retenus. Quelques exemples explicitant ce cas sont présentés ci-dessous :

- A. *Le Blanc appelle la police pour jeter le petit homme su train.*
- *Quel est le rapport exprimé dans cette phrase ?*
  - *Récris-le d'une autre manière.*

<sup>113</sup> Nous ne pouvons à ce niveau généraliser les résultats étant donné que les activités de vocabulaire qui évaluent souvent les connaissances sont parfois liées aux supports proposés. C'est le cas des consignes qui exigent l'identification de mots de nature précise ou le repérage d'éléments précis des textes retenus.

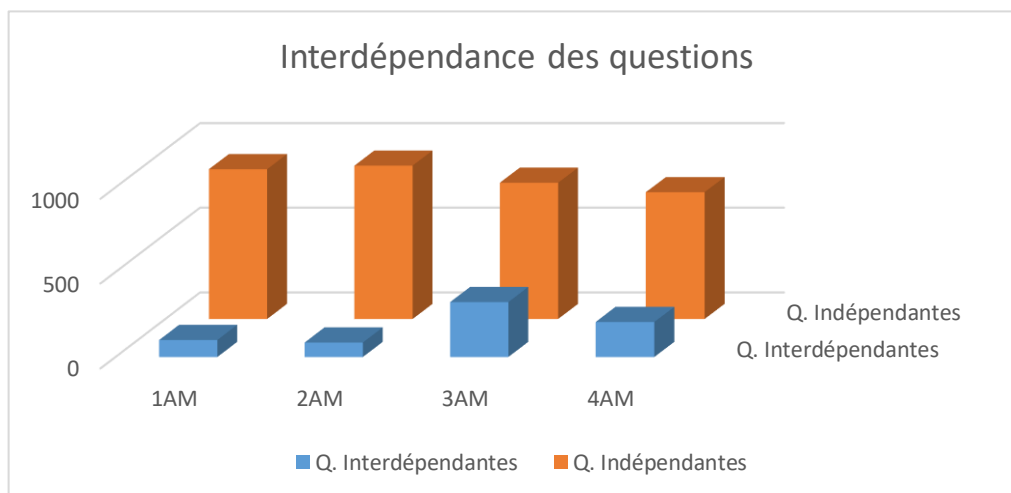


- B. Donne la nature des propositions subordonnées soulignées :
- On affirmait que la diversité unifie une patrie.
  - Le racisme pousse à la violence si bien qu'on doit le combattre.
- C. Mets les verbes entre parenthèses au temps et au mode qui convient :
- Il faut que tu (accepter) la différence.
  - Les trois amis disent que ce comportement (être) injuste.
- D. Réécris les phrases et écris le bon homophone :
- Dans (quel, quelle) région trouve-t (on, ont) l'ours blanc ?
  - Des lois contre la chasse sont sévères, (mai, mais, mets) (c'est, ces, ses) l'homme qui ne les respecte pas.
- E. Transforme la phrase suivante au discours indirect :
- « Je suis très bien à ma place et je n'en bougerai pas », répondit l'homme.

Les consignes ci-dessus ne nécessitent pas le recours aux supports sélectionnés étant donné que les apprenants appliquent ce qui est appris dans des phrases indépendantes qui ne sont pas toujours extraites de ces supports. Vu que ces activités qui évaluent les ressources sont souvent isolées des supports sélectionnés, pourquoi les enseignants continuent-ils à les glisser dans la partie des épreuves consacrée à la compréhension ? Ne faut-il pas les présenter indépendamment des items évaluant la compréhension ?

Cet indicateur, dont les résultats évidents ne sont pas présentés quantitativement, montre que certaines questions que la partie compréhension intègre ne sont pas destinées à évaluer ladite compétence ; ce qui crée des problèmes de validité étant donné que ces questions-ci ne présentent pas de liens étroits aux supports retenus ; d'où la nécessité de leur consacrer une partie autonome.

Le dernier indicateur lié aux questions évaluant la compréhension et les ressources se rapporte à l'interdépendance des questions formulées par les enseignants, déconseillée en évaluation. Cette interdépendance existe-t-elle dans les épreuves constituant notre corpus ? Le graphique suivant présente les résultats dont l'analyse répondra à la question posée.



**Graphique 27: Questions interdépendantes vs questions indépendantes**

Il est incontestable que le taux des questions interdépendantes est nettement faible en comparaison avec les questions indépendantes qui sont dominantes notamment en 1AM et en 2AM. Pour évaluer ces niveaux, les enseignants se contentent de proposer des questions autonomes de façon à évaluer une seule fois la maîtrise de tel ou tel élément. Voici quelques exemples qui illustrent ce premier cas :

- Trouve dans le texte le synonyme de « effet ».
- Propose une solution au problème posé dans le texte.
  - Mets la phrase suivante à la voix passive :  
« L'association des parents d'élèves a tiré la sonnette d'alarme ».

Pour les niveaux les plus avancés du cycle, qui privilégient également les questions indépendantes, ils introduisent certaines questions liées pour mieux vérifier la maîtrise des éléments évalués : une seconde réponse permet aux enseignants de s'assurer si une première réponse correcte est réellement due à une maîtrise de ce qui est appris. Mentionnons que ces questions s'imposent parfois par la nature des supports textuels retenus dans les programmes scolaires, tels que les articles de presse, dont les informations soudées expliquent le recours à des questions liées.

Cette interdépendance des questions est, cependant, déconseillée car la première réponse, si elle est incorrecte, affecte la correction des réponses qui la suivent. Examinons le tableau suivant récurrent dans les épreuves de 3AM.

<i>Quoi ?</i>	<i>Qui ?</i>	<i>Où ?</i>	<i>Quand ?</i>	<i>Comment ?</i>

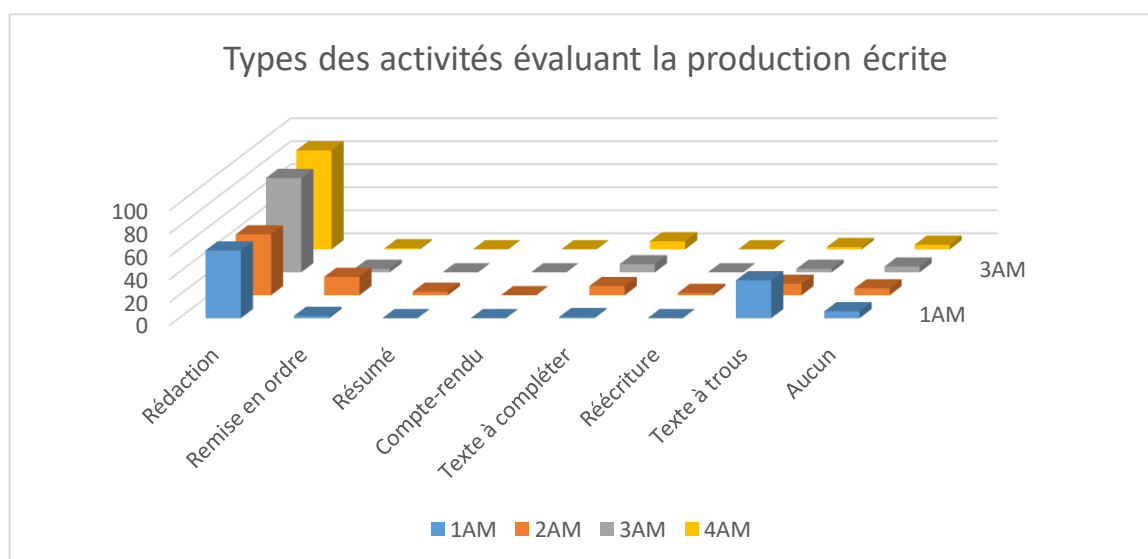
Dans ce tableau, toutes les réponses aux questions posées dépendent de la première question qui renvoie à l'évènement rapporté. Si les apprenants réussissent à repérer cet évènement, ils pourront en déterminer les circonstances. Cette interdépendance, qui conditionne en principe la correction des réponses par celle de la première question, est parfois négligée étant donné que les enseignants –lors de la correction- les traitent comme des questions autonomes pour préserver les chances de leurs apprenants. Il en est de même pour les questions évaluant les ressources -qui en dépit de leur interdépendance- sont corrigées indépendamment les unes des autres. Ainsi, dans les questions ci-dessous, chaque question est notée sans tenir compte de celle qui la précède ou qui la suit :

- Les cigognes percutent les fils électriques de sorte qu'elles meurent.
  - a. Quel est le rapport exprimé dans cette phrase ?
  - b. Exprime-le d'une autre manière.
- Dès que je sais lire et écrire, je peux faire mon travail seul.
  - a. Quel est le rapport exprimé dans cette phrase ?
  - b. L'action de la principale exprime : simultanéité, postériorité, antériorité.

Ceci signifie que cette interdépendance sémantique affecte peu les résultats obtenus par les apprenants étant donné qu'une indépendance correctrice est souvent adoptée par les enseignants du cycle.

Jusqu'ici, les indicateurs quantitativement et qualitativement représentés sont liés aux outils de mesure évaluant la partie Compréhension, qui n'est pas la seule composante évaluée dans les épreuves collectées. Comme ces épreuves sont constituées d'une deuxième partie (Production écrite), intitulée également *Situation d'intégration*<sup>114</sup> dans certaines épreuves recueillies, évaluée à travers des situations, nous lui consacrerons les éléments qui suivent en proposant des indicateurs appropriés. Ces indicateurs-ci apporteront-ils des preuves sur la validité de ces situations introduites pour évaluer la compétence expressive des apprenants ? La réponse à cette question ne pourrait être formulée sans une analyse appuyée sur les données ci-dessous représentées puis analysées.

Le premier indicateur lié à la deuxième partie des épreuves concerne le type des activités que les enseignants sélectionnent en vue d'évaluer la compétence expressive de leurs apprenants. Les résultats collectés sont représentés dans le graphique suivant :



**Graphique 28: Types évaluant la compétence expressive dans les épreuves analysées**

<sup>114</sup> Cette appellation de situation d'intégration que quelques épreuves analysées préfèrent pour insister sur la conformité des épreuves conçues aux exigences révèle néanmoins une certaine confusion entre situation d'intégration et situation d'évaluation en dépit du lien étroit entre les deux. Rappelons que la première est proposée à des fins d'apprentissage (dans une situation d'enseignement/apprentissage) alors que la seconde est introduite à des fins d'évaluation : évaluer la capacité des apprenants à transférer ce qui est appris dans une situation nouvelle.

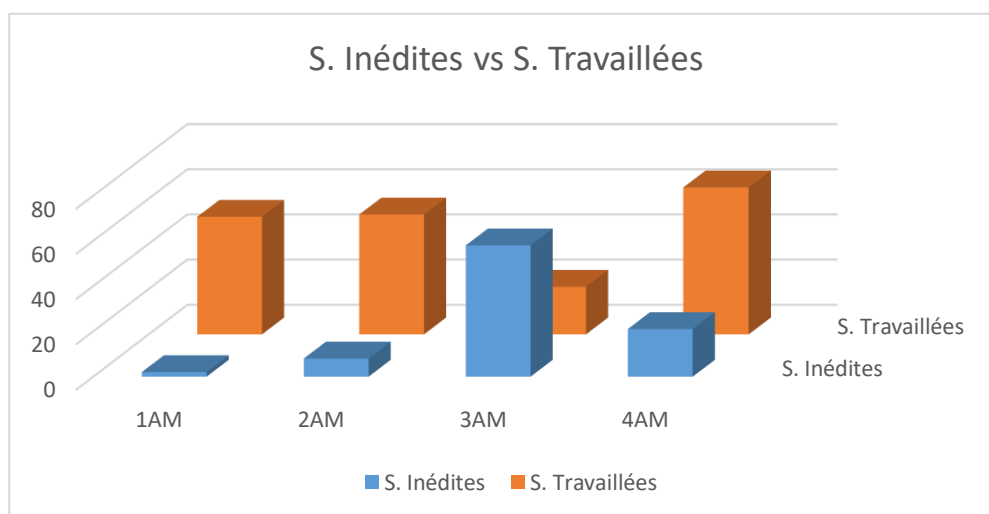
Il est clair que le type dominant dans tous les niveaux, notamment les plus avancés du cycle (3AM et 4AM), est la rédaction, ce qui est nécessaire pour une évaluation de la compétence expressive. Cette activité aboutit, en effet, à des productions concrètes que les enseignants corrigent pour inférer des jugements de compétence. Cette dominance n'empêche pas le recours dans quelques épreuves à des activités comme les textes à compléter, où il est souvent demandé aux apprenants de compléter une partie manquante du texte (chapeau, introduction, développement, conclusion...); ce qui correspond au « texte à gouffre », expression introduite par Bouchard et reprise par Cuq et Gruca (2005).

Cependant, les activités rédactionnelles sont moins privilégiées en 1AM et en 2AM : les enseignants recourent en effet à d'autres activités jugées plus adaptées aux niveaux des apprenants évalués. Il s'agit des *textes à compléter*, des *textes à trous* et des *activités de remise en ordre*. S'agissant des textes à compléter, les enseignants croient atténuer la difficulté des tâches proposées en demandant à leurs apprenants de ne rédiger qu'une partie des textes. Pour les textes à trous, ce sont des activités qui concurrencent les activités rédactionnelles en 1AM où le niveau des apprenants contraint souvent les évaluateurs à proposer des activités jugées plus faciles. Quant aux activités de remise en ordre, elles sont plus introduites en 2AM étant donné que certains enseignants, jugeant leurs apprenants incapables de rédiger des textes narratifs (type textuel retenu pour ce niveau scolaire), préfèrent leur demander de rétablir l'ordre d'une liste d'évènements listés en désordre.

Contrairement aux rédactions qui aboutissent à des productions concrètes permettant l'évaluation de la compétence expressive, les activités ci-dessus évoquées réinvestissent ce qui est proposé sans pour autant parvenir à une production concrète. Seraient-elles valides pour évaluer cette compétence ? Selon la typologie de Cuq et Gruca (2005) qui opposent les activités appropriées à la compréhension à celles adaptées à la production, *les textes à trous* et *les activités de remises en ordre* figurent parmi les activités de la première catégorie, qui évaluent la compréhension : la restitution requiert en effet une compréhension des textes à restructurer. En les introduisant pour évaluer la compétence expressive, les enseignants recourent à des outils évaluatifs non valides. En effet, pour remplir cette qualité, les outils doivent évaluer réellement les compétences qu'ils prétendent évaluer, ce qui n'est pas le cas des deux activités en question. Les *textes à compléter*, ou textes à gouffres, sont cependant rangés par les mêmes auteurs dans les activités de la deuxième catégorie : ces activités permettent en effet d'évaluer la compétence expressive ; ce qui les rend valides en évaluation. Mentionnons un avantage indéniable lié au

recours à ces activités : une évaluation indissociable des deux compétences (compréhension et expression) exigée dans le cadre de l'APC.

Seules les situations proposant des rédactions seront analysées dans les éléments suivants étant donné que les indicateurs retenus visent à en vérifier la validité. Nous vérifions notamment si ces situations sont inédites (nouvelles) et complètes (contenant toutes les composantes essentielles). Le caractère inédit des situations qui s'applique aux situations qui ne sont pas identiques à celles que proposent les manuels scolaires du cycle, souvent traitées et travaillées par les enseignants en classe, est représenté dans le graphique qui suit :



**Graphique 29: Les situations inédites dans le corpus analysé**

Il est clair que les enseignants privilégient les situations déjà travaillées au cours des séquences pédagogiques aux situations inédites dans les épreuves évaluatives conçues même dans les épreuves de 4AM. Le recours à ces situations est certainement peu utile étant donné que les enseignants ne peuvent plus collecter des informations satisfaisantes sur les compétences de leurs apprenants : les situations déjà résolues soumises aux apprenants se transforment, en effet, en des activités de restitution où la capacité de mémorisation est la seule aptitude évaluée. Le recours à ces situations consommées par les enseignants du cycle s'explique par leur souci de proposer des situations adaptées au niveau de leurs apprenants qui se trouvent dépourvus face à des situations inédites.

Voici quelques situations proposées dans les épreuves analysées :

- A. *A l'occasion de la journée mondiale de l'environnement, tu parles à tes camarades sur l'importance d'un environnement sain et sans pollution.*
  - *Rédige un court texte dans lequel tu expliques les conséquences de la pollution et l'importance de protéger notre planète.*

- B. *Ton ami n'aime pas le sport. Il pense que c'est dangereux et qu'on peut se blesser facilement en le pratiquant. Tu n'es pas d'accord avec lui. Tu décides de lui parler des avantages du sport.*
- *Rédige un texte explicatif qui vante les mérites du sport.*
- C. *Dans un paragraphe de 5 à 6 lignes, je raconte la fable du Corbeau et du Renard.*
- D. *Rédige un texte dans lequel tu racontes la fable du lièvre et de la tortue.*

De tous les niveaux, les épreuves de 3AM introduisent plus de situations inédites que les autres. Ceci est dû à la particularité de certains contenus programmés (les articles de presse, les biographies...) qui permettent aux enseignants de proposer aux apprenants des situations évaluatives nouvelles, mais similaires dans leur présentation aux situations d'intégration. Ainsi, pour que des articles de presse soient rédigés, la grille informative extraite de la situation ci-dessous est souvent proposée :

*A partir de la grille suivante, rapporte l'évènement insolite dans une brève de quelques lignes.*

<i>Qui ?</i>	<i>Quoi ?</i>	<i>Où ?</i>	<i>Quand ?</i>	<i>Pourquoi ?</i>	<i>Comment ?</i>
<i>Un étudiant de droit de la faculté de Bordeaux.</i>	<i>Le tribunal (condamner) le tricheur à trois ans de prison ferme.</i>	<i>Bordeaux, France.</i>	<i>La semaine dernière.</i>	<i>Il (tricher) aux examens officiels.</i>	<i>Trois de ses amis (participer) à cette fraude (cette triche).</i>

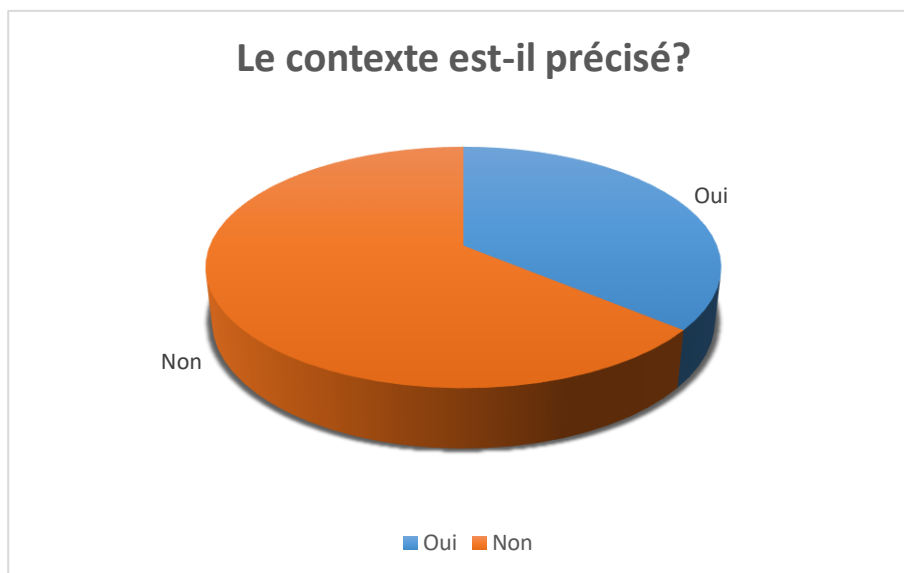
La nouveauté de ces situations pourrait être objectée si nous tenons compte des caractéristiques des situations inédites qui requièrent une articulation différente des mêmes ressources, une caractéristique qui manque dans les situations de 3AM. Celles-ci mobilisent en effet les mêmes ressources notamment que, dans certaines situations évaluatives, les enseignants se contentent d'apporter des modifications légères aux situations d'intégration travaillées en classes.

Excepté le cas particulier des épreuves de 3AM, le nombre des situations inédites –très peu fréquentes– évolue progressivement dans les niveaux du cycle : ces situations sont en effet plus nombreuses en 4AM. Que des situations inédites sont proposées en tenant compte du niveau des apprenants confirme notre explication du recours fréquent aux situations déjà travaillées : le souci d'adapter les outils de mesure au niveau des évalués. Les enseignants jugent, en effet, que les apprenants de 4AM auraient des connaissances et des compétences qui leur permettraient de résoudre des situations nouvelles, ce qui ne s'applique pas aux apprenants de 1AM.

Les résultats ci-dessus exposés révèlent ainsi que la nouveauté des situations n'est pas une exigence très respectée par les enseignants du cycle moyen, ce qui affecte la validité des épreuves conçues. Nous ne pouvons pas cependant émettre un jugement définitif sans vérifier

une deuxième exigence aussi importante : les situations proposées sont-elles complètes ? En vue de répondre à cette question, nous exposons ci-dessous les résultats des indicateurs retenus dans la grille adoptée (Cf. Annexes) dans notre étude.

- Le contexte est-il précisé dans les situations analysées ?



**Graphique 30: La précision du contexte dans les situations évaluatives**

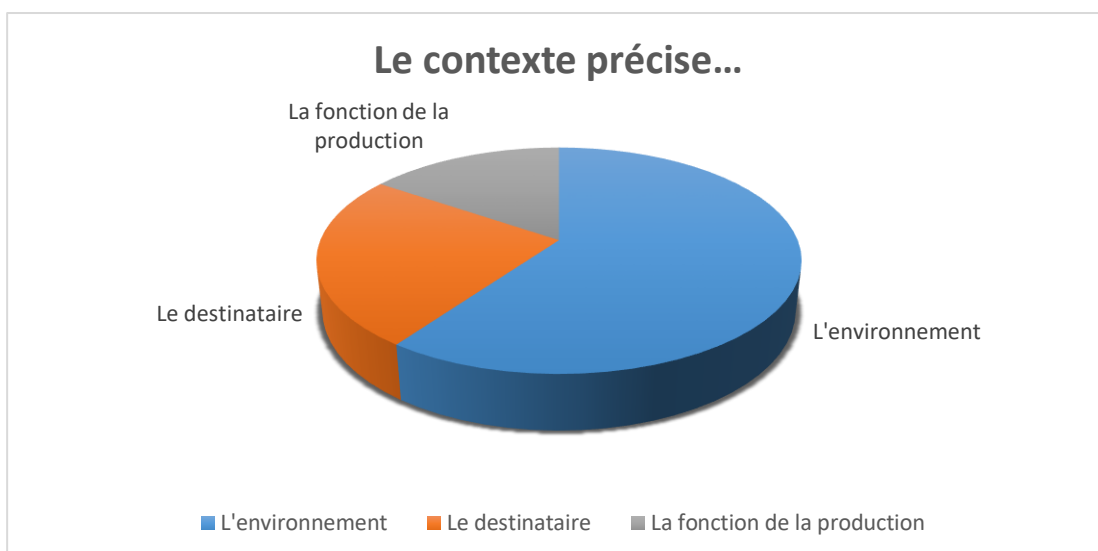
Il est clair que la majorité des situations n'explicitent pas le contexte des tâches proposées aux apprenants du cycle moyen. Dans certaines situations, en effet, les enseignants leur soumettent directement les tâches à réaliser. Ils privilégient ainsi une formulation claire, précise et notamment concise pour éviter des précisions supplémentaires qui pourraient écartier les évalués du thème à traiter. Les situations suivantes illustrent ce cas-ci :

- *Rédige un fait divers à partir des informations suivantes.*
  - *Rédige le conte de Cendrillon.*
- *Rédige avec tes propres mots une fable de ton choix.*

Dans les situations où le contexte est explicité, des précisions sur les circonstances, la fonction, les destinataires...des productions sont parfois fournies par les évaluateurs. Nous retrouvons ces précisions dans la situation ci-dessous destinée à des apprenants de 1AM.

- Ton petit frère joue beaucoup dans le jardin. Il rentre à la maison les mains tout sales.*
- *Ecris un texte pour lui expliquer l'importance de l'hygiène des mains (se laver les mains) et comment y procéder correctement.*

Comme les précisions qui explicitent le contexte sont nombreuses, laquelle est plus fréquente dans les situations analysées ? Les résultats collectés sont répartis sur trois variables : environnement (circonstances, temps et lieu), destinataire et fonction de la production.



**Graphique 31: Répartition des précisions dans les contextes des situations**

Souvent, les contextes précisent l'environnement de la production pour que la situation de communication soit claire aux apprenants. Comme l'environnement, tel que les situations le présentent, ne précise pas tous les paramètres de la situation, la tâche prescrite détermine parfois les destinataires de la production. Il en est de même pour la fonction qui met en valeur la visée des textes à produire par les apprenants (expliquer, convaincre, pousser à agir...). Ces trois éléments se réunissent parfois dans une même situation comme le montre l'exemple suivant :

*Lors de la récréation, tu as remarqué que la majorité de tes camarades mange des bonbons et du chocolat. Alors, t'as décidé de les avertir des dangers de cette mauvaise alimentation.*

- *Rédige un paragraphe dans lequel tu expliqueras à tes camarades l'importance d'une alimentation équilibrée.*

Ceci ne s'applique pas à toutes les situations étant donné que nombreuses sont celles qui ne citent qu'un seul élément. Voici quelques exemples illustrant ce cas-ci :

- *Ecris un texte d'une quarantaine de mots pour expliquer à tes camarades les principales règles du sport que tu pratiques.*
- *Rédige un texte de sept lignes pour expliquer les dangers d'une mauvaise alimentation.*

En dépit de leur importance, ces éléments du contexte sont souvent omis des situations évaluatives conçues par les enseignants qui préfèrent des situations annonçant directement la tâche à réaliser :

- *Raconte une fable de ton choix.*
- *Explique l'importance de se laver les mains.*

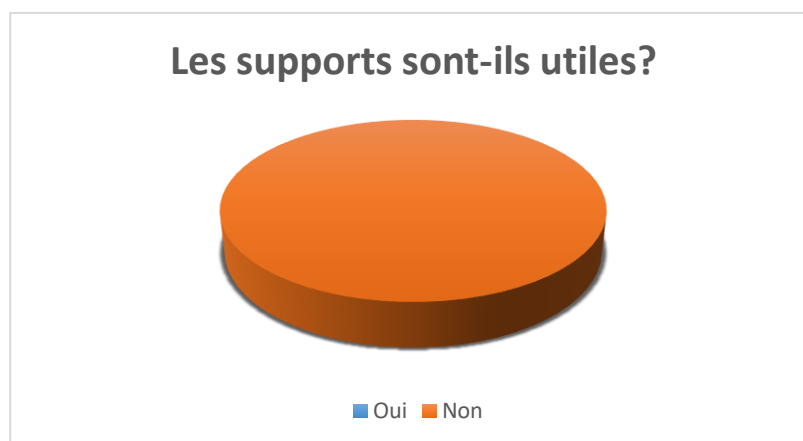
Le contexte ne constitue pas le seul élément que de nombreuses situations évaluatives du cycle moyen omettent. Les supports, constituants indispensables, sont également négligés par les enseignants. C'est ce que montre le graphique suivant :





**Graphique 32: Les constituants des situations analysées**

Comme les indicateurs antérieurement exposés l'ont mentionné, peu de situations incluent des supports, qui consistent souvent en des illustrations peu utiles à la résolution des tâches proposées.

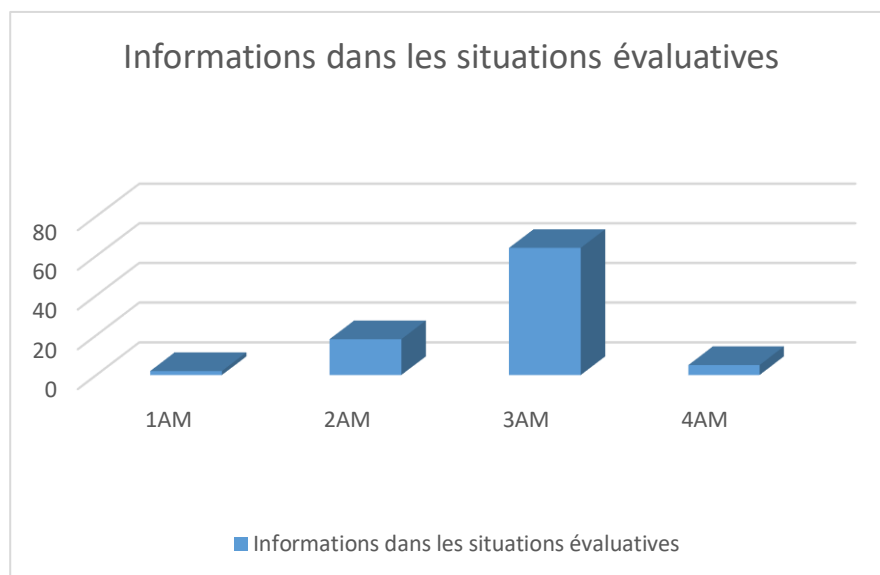


**Graphique 33: L'utilité des supports intégrés dans les situations évaluatives**

De prime abord, les résultats liés à cet indicateur semblent subjectifs étant donné que cette utilité ne pourrait être aisément mesurée. Cependant, en associant cette utilité à la capacité des supports sélectionnés de fournir des renseignements utiles à la résolution des situations-problèmes, la tâche devient plus aisée. Pour les supports que les enseignants sélectionnent dans leurs situations d'évaluation analysées, ils ne fournissent pas aux apprenants des informations pertinentes qui les aident à réaliser les tâches prescrites, ce qui est dû à deux facteurs essentiels. Le premier est lié à la nature de ces supports qui consistent en des illustrations représentant souvent les thèmes traités, les personnalités présentées, les personnages mis en scène..., ce qui permet aux apprenants d'identifier les thèmes de leurs productions. Le second facteur concerne les enseignants qui pourraient concevoir ces situations exemptes de supports utiles intentionnellement ou involontairement. L'omission intentionnelle des supports s'applique aux

situations qui ne sont pas inédites étant donné que les apprenants ont les informations nécessaires pour les résoudre ; et à celles conçues par des enseignants réticents au changement. Quant à l'omission involontaire, elle révèle une certaine ignorance des caractéristiques des situations évaluatives et des exigences de l'évaluation des compétences de la part des enseignants. Mais, elle pourrait être due également aux pratiques habituelles des enseignants qui entravent l'intégration des nouvelles exigences. Mentionnons que ces changements pourraient être rejetés par les apprenants habitués à des pratiques et des activités particulières en évaluation.

De cette manière, les situations évaluatives que les enseignants proposent au cycle moyen s'avèrent parfois incomplètes, ce qui les rend non valides en évaluation du FLE. Si les enseignants omettent souvent les supports de leurs situations, ils proposent parfois des informations que les apprenants doivent nécessairement investir pour la rédaction des différents types textuels. Contrairement aux supports qui alternent des renseignements utiles et inutiles, ces informations-ci, toujours utiles, sont des contraintes imposées aux apprenants qui les intègrent dans leurs rédactions. Ce type d'activités d'écriture (matrices de textes) cité par Cuq et Gruca (2005) consistent à produire des textes à partir de contraintes précises. Il est important de préciser que ces informations ne sont notamment fournies aux apprenants de 3AM pour que les situations qui leur sont destinées soient inédites.



**Graphique 34: Les informations dans les situations d'évaluation du cycle moyen**

Comme le graphique ci-dessus le montre, ce sont les situations (notamment inédites) de 3AM qui soumettent aux apprenants des informations qui concernent les personnalités à présenter (*nom, prénom, date et lieu de naissance, profession, œuvres...*), les événements à

rapporter (*date, lieu, causes, conséquences...*)... Ces informations sont moins fréquentes en 2AM où les enseignants proposent rarement aux apprenants des indications sur les histoires à raconter, les personnages à décrire... En 1AM et en 4AM, les apprenants s'appuient sur leurs prérequis pour résoudre les situations proposées étant donné que celles-ci sont dépourvues de supports, ce qui les non-conformes aux exigences de l'APC.

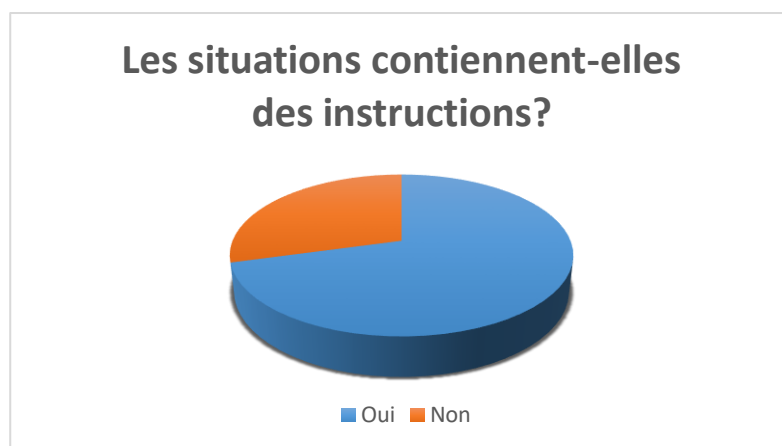
Un autre constituant est à vérifier pour que les résultats de notre étude soient plus fondés. Il s'agit du nombre des tâches que les enseignants proposent pour évaluer les compétences de leurs apprenants. Rappelons que trois tâches au moins doivent être présentées aux apprenants et que deux de ces tâches doivent être réussies pour émettre un jugement de compétence (règle des 2/3). Les situations d'évaluation analysées respectent-elle cette règle ? Proposent-elles aux apprenants le nombre de tâches requis ? Le graphique ci-dessous répond à cette question.



**Graphique 35: Le respect de la règle des 2/3 dans les situations du corpus**

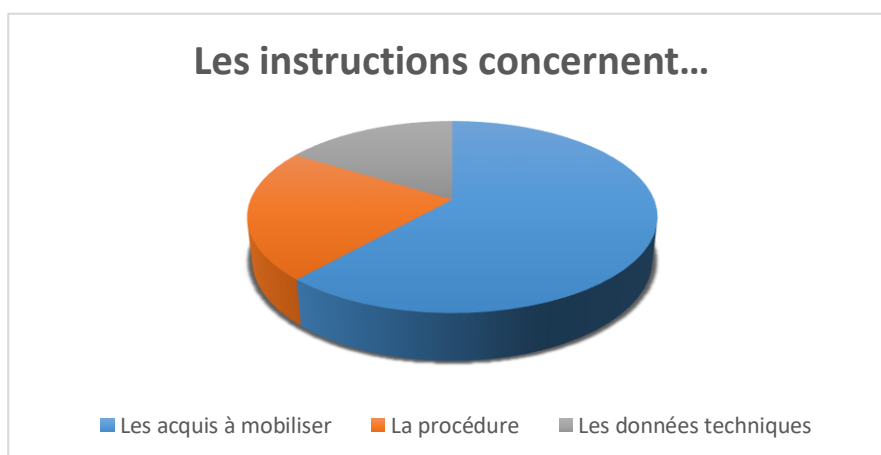
Toutes les situations constituant notre corpus se contentent d'une tâche unique à réaliser par les apprenants. Ainsi, ceux-ci sont reconnus compétents en réussissant cette tâche, et incompétents en cas d'échec ; ce qui signifie que la règle des 2/3, indispensable à la validité de l'évaluation dans une perspective de compétences, est écartée.

Les instructions de réalisation constituent la dernière composante que nous étudierons. Contrairement aux autres composantes antérieurement abordées, elles sont fréquentes dans les situations d'évaluation, comme le graphique ci-dessous l'illustre. Nous les retrouvons même dans des situations incomplètes qui ne précisent que la tâche à réaliser. Que ces instructions précisent-elles ?



**Graphique 36: Les instructions dans les situations d'évaluation**

Les résultats qui répondent à la question précédente sont répartis sur les trois variables retenues : les acquis à mobiliser, la procédure de réalisation et les données techniques.



**Graphique 37: La nature des instructions que les situations contiennent**

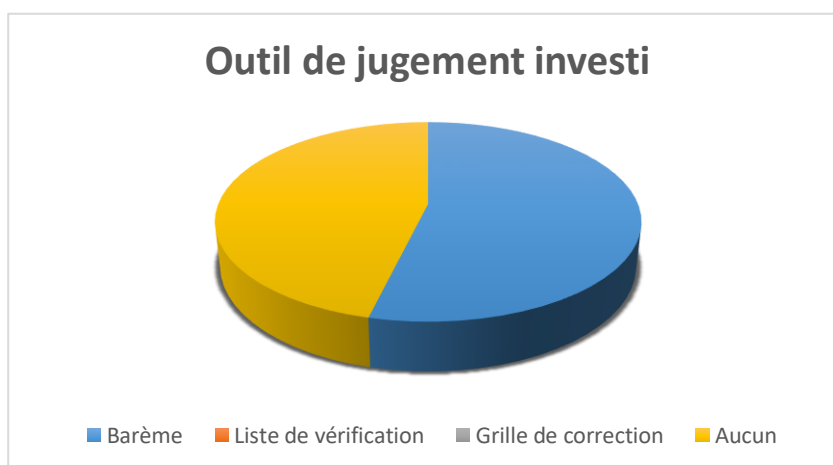
Souvent, les instructions (dits critères de réussite dans les épreuves collectées) déterminent les acquis à mobiliser par les apprenants, autrement dit les ressources. Ces acquis englobent le vocabulaire à investir (*banque de mots*) ; les temps de conjugaison à employer (*utilise le passé simple, mets les verbes au présent de l'indicatif...*) ; le type et la forme des phrases (*propose une question, introduis une phrase à la forme passive...*) ... Quant à la procédure, elle précise la progression des éléments de la production (*rédige une introduction, propose trois arguments, illustre-les par des exemples...*). Enfin, les données techniques –qui exigent rarement une disposition particulière de la production (*présente ton article en colonnes*)- concernent essentiellement la longueur des textes à produire (*rédige un texte de cinq lignes...*). Une présentation aussi détaillée des instructions de réalisation ne passe pas inaperçue étant donné que, contrairement aux situations d'intégration, les apprenants doivent eux-mêmes déterminer les ressources à mobiliser pour résoudre les problèmes soulevés. Ces instructions

trahissent ainsi une certaine implication des enseignants qui s'en servent pour transmettre des recommandations et orienter les apprenants.

Nombreux, les indicateurs liés aux outils de mesure contestent sérieusement la validité des épreuves conçues pour évaluer les compétences des collégiens algériens en FLE. Nous exposerons dans ce qui suit ceux relatifs aux outils de jugement investis pour déterminer à quel point ils sont adaptés aux exigences de l'APC.

#### 2.4. Les outils de jugement

A travers le premier indicateur, nous déterminerons les outils de jugement investis dans notre corpus. Les résultats sont répartis sur quatre variables comme le présente le graphique suivant :

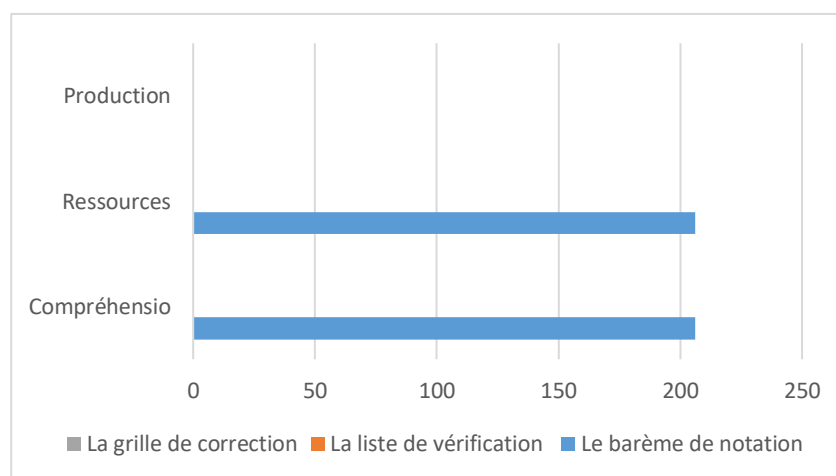


**Graphique 38: Les outils de jugement investis dans les épreuves du FLE**

Il est clair que le seul outil que les enseignants investissent pour évaluer les performances de leurs apprenants est le barème de notation. Ce choix est justifié pour évaluer la compréhension et les ressources, où des items sont sélectionnés, mais il ne convient pas pour corriger les situations d'évaluation évaluant la production. Cependant, la répartition des notes ne figure pas toujours dans les épreuves analysées : le barème est en effet omis dans 46% de ces épreuves. Pour quelles raisons n'est-il pas mentionné ?

Il convient de citer deux tendances qui contribuent à répondre à notre question. La première est empruntée par les enseignants qui préfèrent mentionner le barème sur le sujet pour garantir une certaine équité et éviter toute subjectivité. Ces avantages anticipent les dérives de la deuxième tendance où les enseignants adaptent leurs barèmes aux réponses de leurs apprenants (les copies des apprenants sont feuilletées avant la conception du barème). Comme cet outil de jugement ne pourrait pas être individualisé, son adaptation est toujours injuste pour

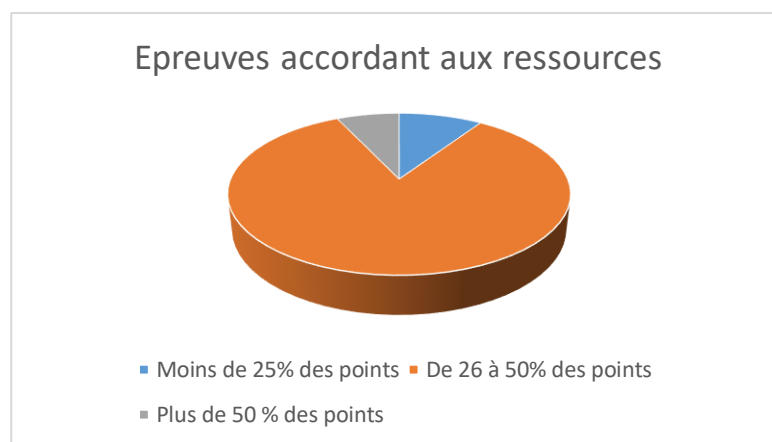
une catégorie d'apprenants. Comme les barèmes sont conçus en fonction des réponses formulées par les apprenants, nous pouvons considérer que la répartition des notes par les enseignants est aléatoire, ne dépendant pas de critères précis (importance du contenu). La validité de cet outil se trouve ainsi altérée en raison de sa manipulation par des pratiques enseignantes subjectives, mais également à cause de la nature des outils de mesure investis. En effet, les épreuves que nous avons analysées investissent deux outils : des items et des situations, qui requièrent deux outils de jugement distincts : le barème et la grille de correction. Le barème, étant le seul outil de jugement introduit, est-il appliqué également aux situations ? Quelles sont les parties de cet outil évalué-t-il ?



**Graphique 39: Les parties évaluées par le barème de notation**

Le barème de notation évalue la compréhension et la maîtrise des ressources. Mais, il n'évalue pas les situations évaluatives évaluant la production. D'ailleurs, nous ne repérons aucun outil évaluant cette partie dans les épreuves analysées. De cette manière, les outils de jugement appliqués à la correction des outils de mesure introduits ne sont pas toujours communiqués aux apprenants, ce qui percute les recommandations des auteurs –à l'instar de Côté, Tardif et Munn (2011)- qui considèrent nécessaire la communication des critères de correction aux évalués.

Comme le barème de notation, outil privilégié, indique la répartition des notes pour les questions évaluant les ressources, nous examinons en dernier lieu si le taux de points qui leur est accordé ne dépasse pas 25%, comme le préconise Gérard (2009).



**Graphique 40: Le taux accordé aux ressources dans les épreuves**

Le graphique précédent recense les épreuves qui accordent aux ressources (contenus linguistiques) moins de 25% des points révèle que leur nombre est restreint, ce qui ne correspond pas aux instructions de l'APC. En effet, la quasi-totalité des épreuves constituant notre corpus leur attribue plus que le taux fixé. Ce qui est surprenant est que certaines évaluations, en dépit de leur nombre restreint, leur accorde plus de 50%. De tels résultats révèlent l'importance que requiert la maîtrise des contenus linguistiques aux yeux des enseignants qui associent ainsi la compétence de leurs apprenants à la maîtrise des savoirs et savoir-faire linguistiques.

Notre analyse des résultats exposés ci-dessus révèle alors que les épreuves conçues par les enseignants du cycle moyen -pour évaluer les compétences écrites de leurs apprenants en FLE- présentent plusieurs problèmes de validité étant donné qu'elles ne respectent pas toutes les exigences de l'APC. Parmi ces problèmes, notons :

- L'évaluation séparée des compétences ;
- Le recours aux items pour évaluer la compréhension ;
- L'introduction de situations d'évaluation incomplètes et déjà travaillées ;
- Le recours au barème de notation, inadapté à certains outils de mesure (les situations d'évaluation) ;
- Le taux de points accordé aux ressources...

Ces problèmes sont-ils suffisants pour que les épreuves du FLE au cycle moyen soient déclarées non valides ?

### 3. Discussion

De tous les problèmes de validité soulevés précédemment, un seul indicateur pourrait être contesté : l'évaluation séparée des compétences. Rejetée dans une perspective de

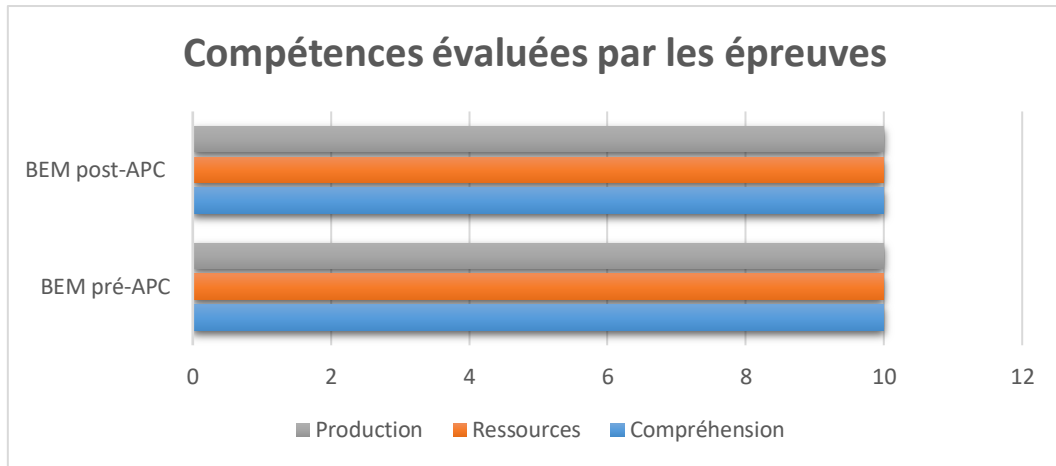
compétences, cette évaluation séparée existe toujours dans les tests les plus prestigieux de FLE dont la validité ne peut être objectée. Les tests DALF, DELF, DILF... illustrent nos propos car la compréhension et la production y sont évaluées séparément en introduisant des outils appropriés. Pour évaluer la compréhension, ces tests privilégient des questionnaires, sur les supports sélectionnés, qui ne nécessitent pas une formulation des réponses (QCM, association, cases à cocher...), ce qui aboutit à des outils valides n'évaluant que la compétence réellement mesurée. Cependant, les outils évaluant la production ne sont pas toujours valides étant donné que la compréhension des consignes est nécessairement mobilisée pour la résolution des tâches prescrites, ce qui biaise la validité des activités proposées. Ainsi, la séparation des compétences, qui ne présente théoriquement aucun risque sur la validité des épreuves, devient impossible en évaluation des langues vu la complémentarité indiscutable des deux compétences. Mentionnons que si ces outils d'évaluation investis par ces tests du CECRL ne correspondent pas aux exigences de l'APC et aux instructions officielles, leurs outils de jugement adoptés (les grilles critériées) pourraient largement inspirer les enseignants du cycle moyen.

En réalité, cette évaluation dissociée des compétences en évaluation du FLE n'est pas récente : les épreuves sommatives conçues dans le cadre de la PPO, qui a précédé l'APC, ont également évalué la compréhension et la production séparément. Cet élément commun constaté entre les épreuves conçues dans les deux pédagogies nous amène à supposer que les pratiques évaluatives n'ont pas réellement évolué pour être conformes aux instructions de la nouvelle approche. Notre hypothèse rejoint la conclusion à laquelle est parvenue Ferhani-Meghraoui suite à une étude des épreuves du BAC. Cette auteure affirme en effet que « *les anciennes pratiques évaluatives perdurent malgré les opérations de formation continue organisées par le MEN au profit des enseignants et des formateurs (enseignants-formateurs et inspecteurs)* » (2019, p. 46). Pour vérifier si cette conclusion s'applique à notre corpus, une comparaison entre les sujets de **BEM pré-APC** à ceux **post-APC**.

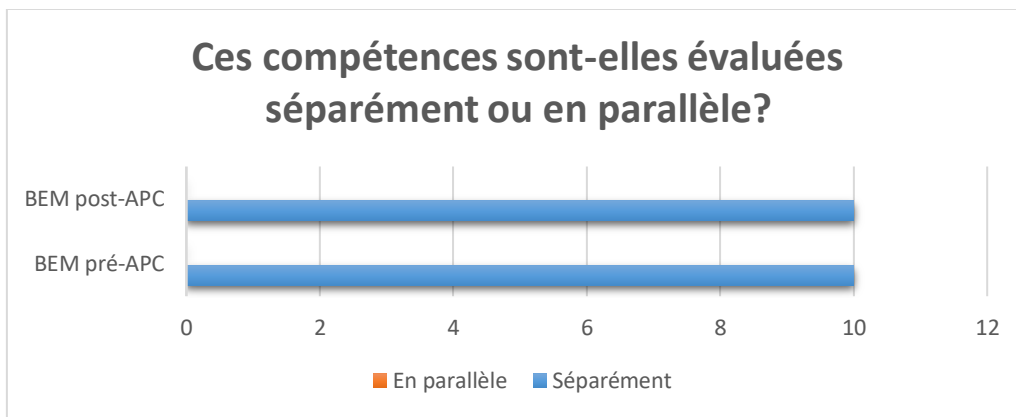
### 3.1. Les compétences évaluées

Les résultats liés à cet élément seront exposés en dessous des représentations graphiques qui suivent :

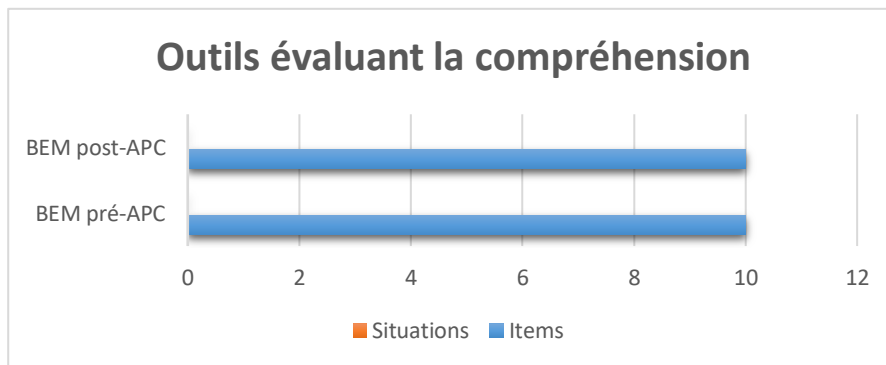




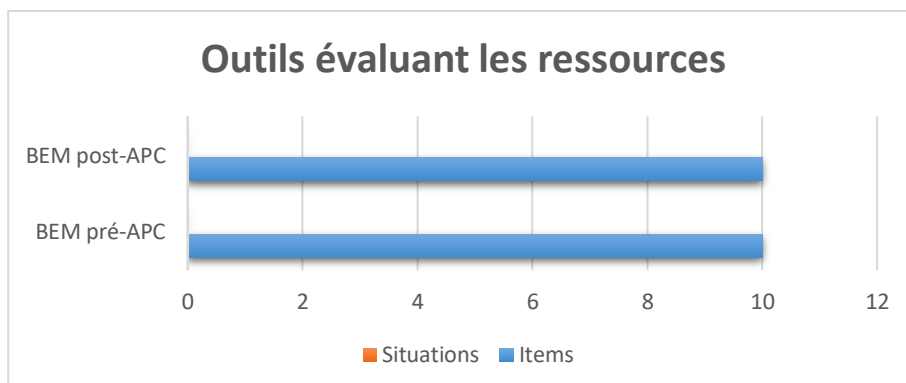
**Graphique 41: Répartition des compétences évaluées dans les épreuves**

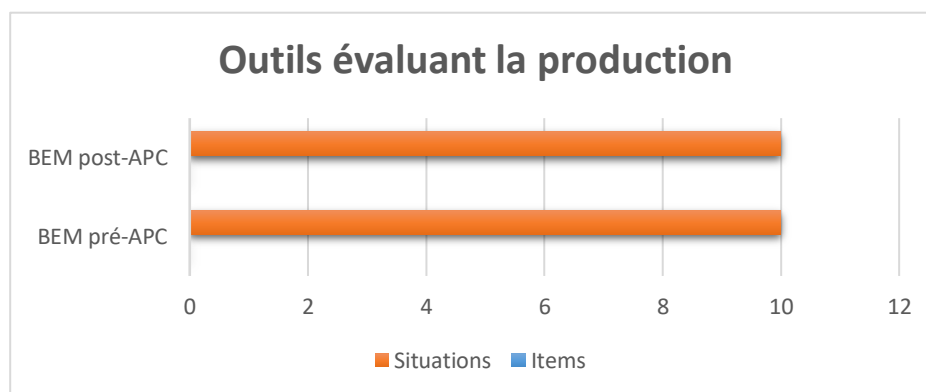


**Graphique 42: Compétences évaluées séparément ou en parallèle**



**Graphique 43: Outils évaluant la compréhension**



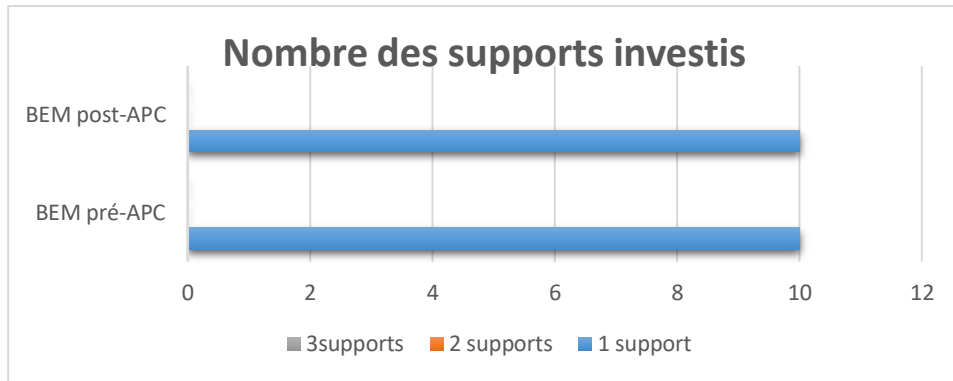
**Graphique 44: Outils évaluant les ressources****Graphique 45: Outils évaluant la production écrite**

Cette première série révèle que les compétences évaluées dans les deux catégories sont identiques ainsi que les outils qui les évaluent : des items sont investis pour évaluer la compréhension et les ressources tandis que des situations évaluent la production. Ces compétences, évaluées séparément, ne sont plus présentées de la même manière : une certaine évolution, en effet, a touché la structure de ces épreuves. Les BEM pré-APC sont constitués de trois parties évaluant chacune une des trois composantes : *la compréhension de l'écrit, le fonctionnement de la langue et la production écrite*. Cette organisation des épreuves disparaît en 2007, date du premier brevet conforme aux exigences de la nouvelle approche. Les épreuves sont réparties dorénavant en deux parties : *la compréhension de l'écrit et la production écrite*. Cette modification, qui annonce manifestement un changement profond des pratiques évaluatives (suppression des questions évaluant les ressources), ne concerne en réalité que la forme des épreuves étant donné que la partie consacrée à la compréhension de l'écrit intègre souvent des questions évaluant les contenus grammaticaux enseignés. Les épreuves post-APC, qui devraient être valides, se trouvent ainsi biaisées étant donné que les items destinés à n'évaluer que la compréhension évaluent une autre composante (contenus linguistiques).

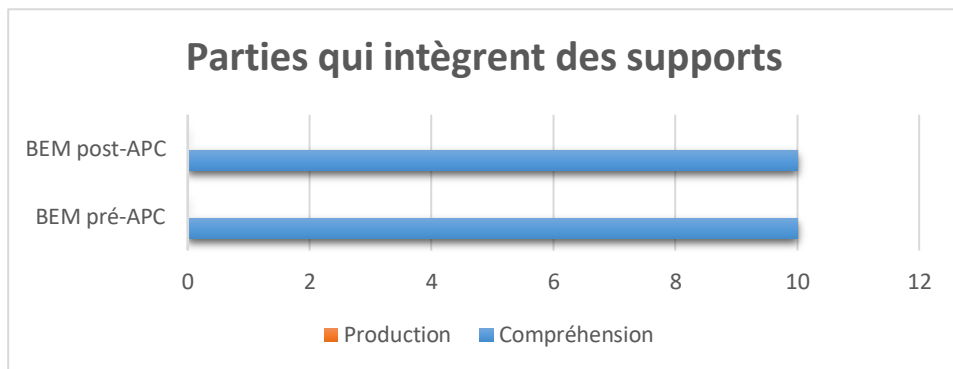
En somme, la réforme du système éducatif algérien qui a introduit une nouvelle approche (APC) a préservé la même structure des épreuves évaluatives où trois composantes (compréhension, outils linguistiques et ressources) sont toujours évaluées séparément ; en adoptant les mêmes outils.

### 3.2. Les supports sélectionnés

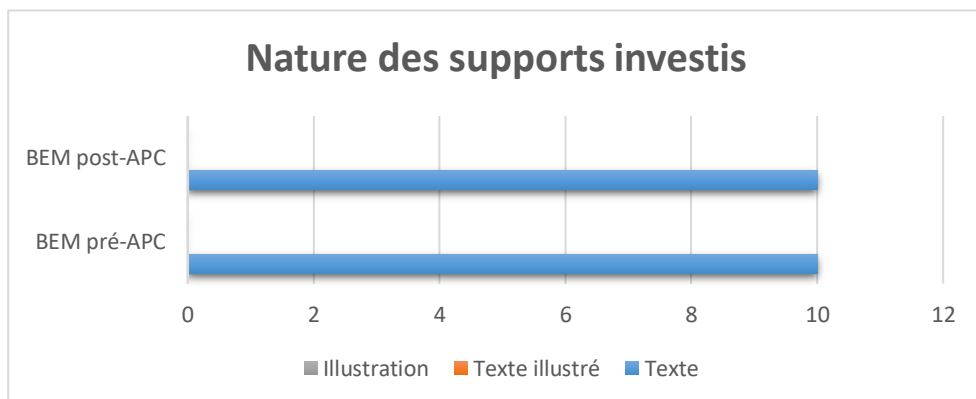
Nous présenterons les résultats relatifs à cette partie dans les graphiques ci-dessous :



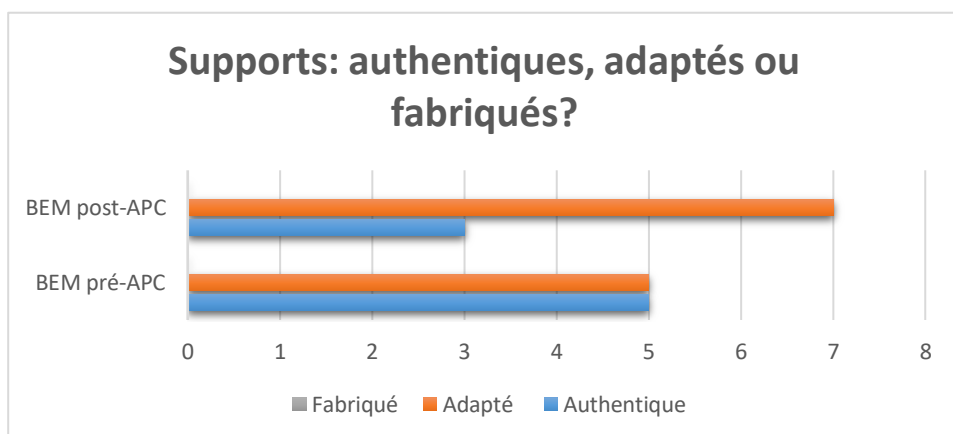
**Graphique 46: Nombre de supports sélectionnés dans les BEM analysés**



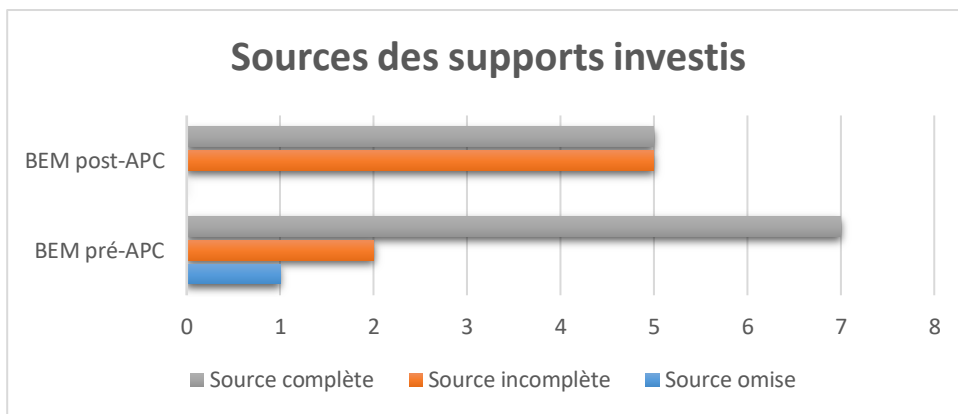
**Graphique 47: Parties qui intègrent des supports**



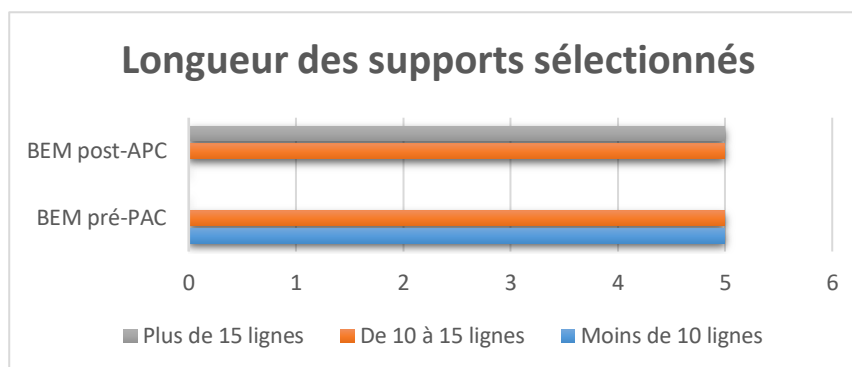
**Graphique 48: Nature des supports investis**



**Graphique 49: Supports des épreuves analysées : authentiques, adaptés ou fabriqués ?**



**Graphique 50: Précisions sur les sources investis**



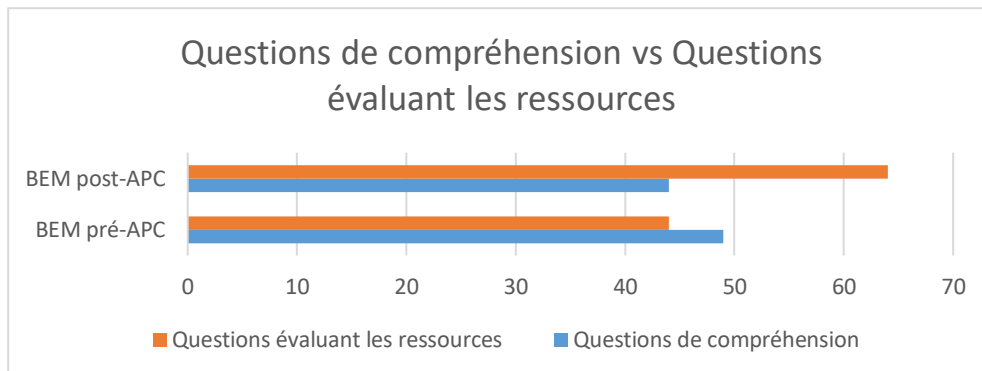
**Graphique 51: Longueur des supports sélectionnés dans les BEM**

Dans les deux catégories de BEM, un support textuel unique est proposé dans la partie qui évalue la compréhension. Si le nombre et la nature de ces supports sont stables, certaines caractéristiques ne le sont plus comme la longueur. En effet, les épreuves ont tendance à recourir actuellement à des textes plus longs contrairement aux BEM pré-APC où des textes très courts (05 lignes) ont été sélectionnés. Cette évolution est peu pertinente dans une perspective de validité étant donné que certains textes courts se caractérisent par une densité informative et une complexité linguistique qui les rendent aussi importants que les textes longs. Bien que cette évolution soit révélatrice d'une sélection plus minutieuse des supports à proposer aux apprenants pour les évaluer en FLE, le recours constaté aux textes adaptés dans les épreuves post-APC ne passe pas inaperçu notamment que les instructions recommandent vivement la sélection de documents authentiques. Ces épreuves où les textes adaptés sont plus fréquents que les supports authentiques en mentionnent toujours les références, contrairement aux BEM pré-APC où la source du texte soumis aux évalués est omise.

En réalité, la comparaison des supports des épreuves de BEM pré-APC à ceux des épreuves post-APC ne révèlent pas de différences significatives ; et n'annoncent pas par-là une certaine évolution des pratiques évaluatives.

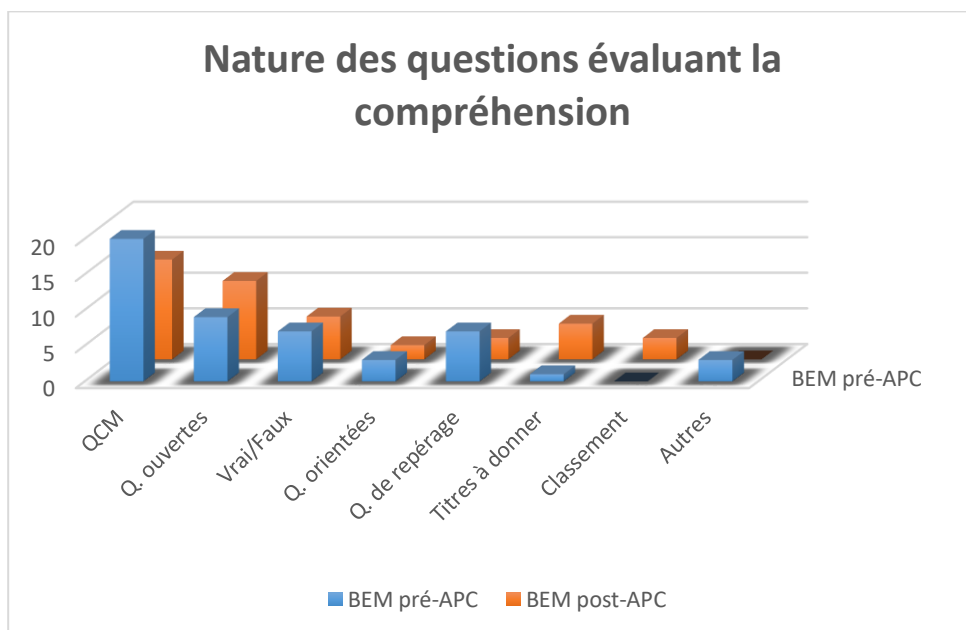
### 3.3. La compréhension et les ressources : outils intégrés

Pour évaluer ces deux composantes, les épreuves de BEM recourent à des items. Leur nombre qui varie d'une épreuve à une autre est représenté dans le graphique suivant :



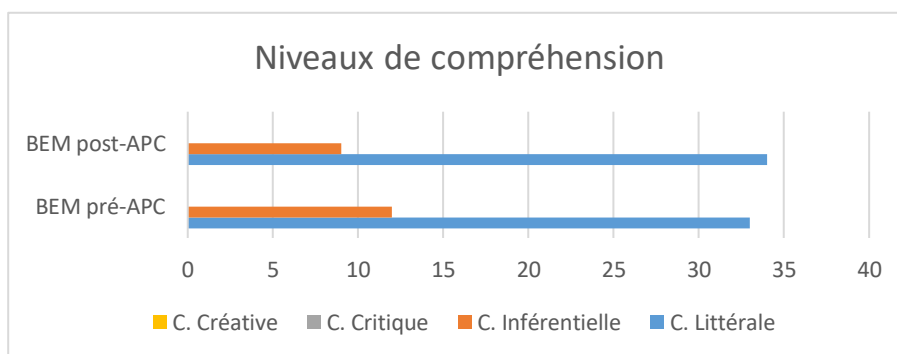
**Graphique 52: Comparaison des questions évaluant la compréhension à celles évaluant les ressources**

Ce graphique, qui montre que les questions sont nombreuses dans les épreuves post-APC, nous permet une comparaison du nombre des questions évaluant la compréhension à celui des questions qui portent sur les outils linguistiques. En effet, il est apparent que les questions qui évaluent les ressources sont plus nombreuses dans les épreuves qui s'inscrivent dans une perspective de compétences, ce qui traduit une priorité accordée à la maîtrise des outils linguistiques par les évaluateurs. Une telle priorité confirme de nouveau que les pratiques évaluatives du cycle moyen – qui ne sont pas conformes aux exigences de l'APC- ont stagné, d'où la nécessité de réformer l'évaluation scolaire. Par quels outils la compréhension et les ressources sont-ils évalués ?



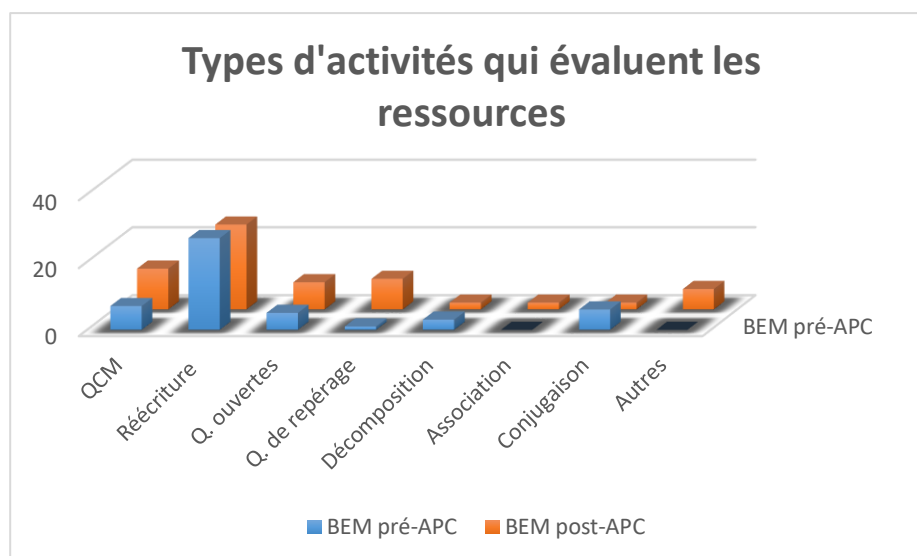
**Graphique 53: Types de questions évaluant la compréhension dans les BEM**

A l'instar des représentations précédentes, celle-ci révèle que les types d'activités évaluant la compréhension sont identiques dans les BEM pré-APC et post-APC. QCM, questions ouvertes, questions Vrai/faux... sont en effet des types récurrents dans les deux catégories. Notons quelques différences, peu significatives, au niveau du nombre de questions relevant de chaque type : QCM plus nombreux dans les BEM pré-APC, questions ouvertes plus nombreuses dans les BEM post-APC... Mentionnons que toutes les questions de compréhension évaluent notamment la compréhension littérale des supports textuels proposés ainsi que la compréhension inférentielle.



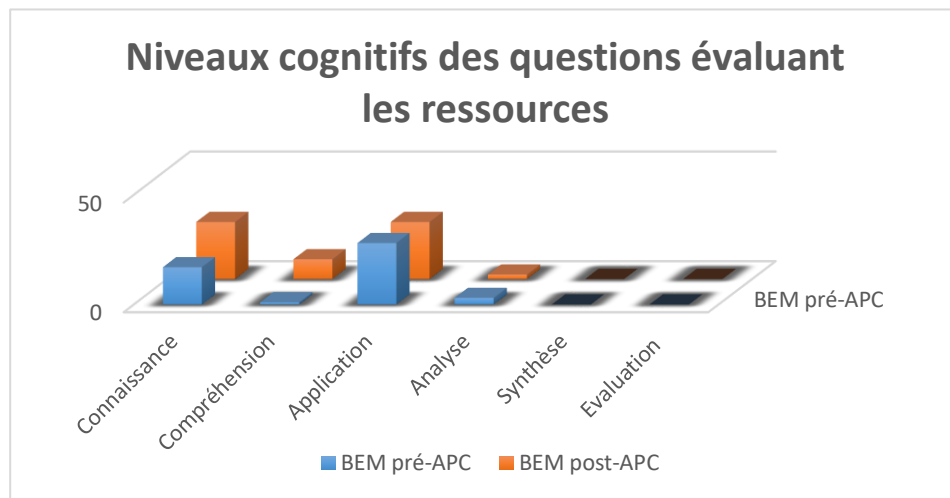
Graphique 54: Niveaux de compréhension évalués

Pour les questions évaluant les ressources, leurs types sont plus nombreux dans les BEM post-APC où la maîtrise des outils linguistiques revêt une importance indéniable. Ceci n'empêche pas le recours fréquent dans toutes les épreuves aux activités de réécriture (*transformation de phrases décontextualisées de la voix active à la voix passive, du discours direct au discours indirect, du singulier au pluriel...*).



Graphique 55: Types des activités évaluant les ressources

Si les types des activités sont quasi-identiques, les niveaux cognitifs visés sont-ils similaires dans les deux catégories d'épreuves ?



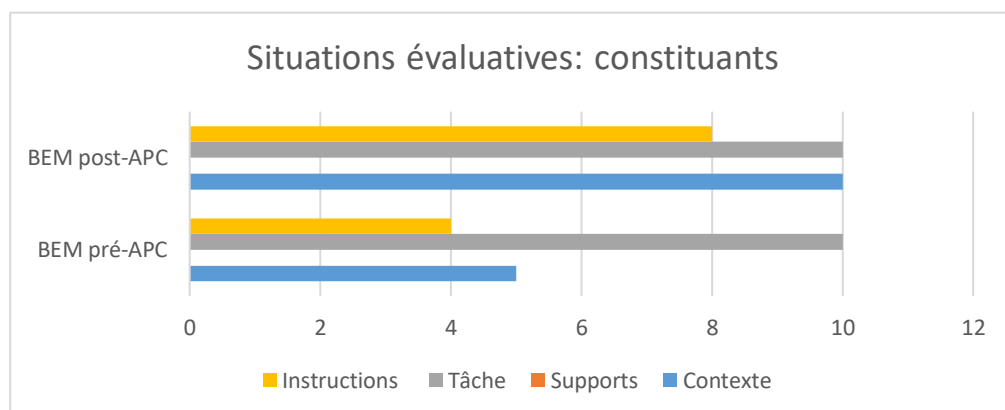
**Graphique 56: Les niveaux cognitifs des questions qui évaluent les ressources**

Ce graphique, comme les précédents, montre que les deux niveaux les plus visés en évaluation des outils linguistiques sont la connaissance et l'application, ce qui révèle que le système éducatif accorde une priorité à la maîtrise des savoirs et savoir-faire même après l'introduction de l'APC.

Le dernier indicateur (que nous ne représentons pas graphiquement) est l'interdépendance des questions proposées aux apprenants. Bien qu'elles soient rares, certaines questions interdépendantes existent dans les deux catégories de BEM pour évaluer les ressources, ce qui ne respecte pas une des exigences de l'APC.

### 3.4. La production écrite : comparaison des situations

Pour confronter les situations évaluatives proposées dans les deux catégories de BEM, nous étudierons leurs composantes :



**Graphique 57: Les éléments constitutifs des situations évaluatives**

Toutes les situations analysées, qui ne proposent qu'une tâche unique et ne respectent pas par-là la règle des 2/3 indispensable à leur validité, ne sont pas complètes étant donné que les supports n'y figurent pas. Les apprenants, dans ce sens, s'appuient sur leurs informations personnelles pour résoudre la tâche proposée ; ce qui défavorise certainement les apprenants peu informés sur les thèmes traités si les situations sont inédites. Si les situations ne le sont pas, elles se transforment en des activités de restitution évaluant plutôt la capacité de mémorisation. Bien que les situations post-APC, dépourvues de supports, ne soient pas complètes, elles accordent plus d'importance au contexte (qui précise notamment les circonstances) et aux instructions (qui concernent essentiellement la longueur des productions à rédiger).

**3.5. Les outils de jugement**

Aucun outil de jugement à adopter par les enseignants pour la correction des réponses des apprenants n'est mentionné dans les BEM analysés, ce qui laisse les apprenants perplexes quant aux critères à investir par leurs évaluateurs. La note attribuée à chacune des parties constituant les épreuves est cependant mentionnée : les BEM pré-APC accordent huit (08) points à la compréhension, six (06) points aux ressources et six (06) points à la production ; alors que les BEM post-APC accordent treize (13) points à la compréhension (qui englobe les ressources) et sept (07) à la production. Vu que la répartition des notes sur les questions proposées n'est pas clairement indiquée, nous ne pouvons déterminer si le taux accordé aux ressources ne dépasse pas le taux requis : 25%.

La confrontation des épreuves de BEM évaluant les compétences des collégiens en FLE avant l'introduction de l'APC à celles qui les ont évalués après son introduction révèle que les évaluations sommatives du FLE n'ont pas évolué pour s'adapter à l'évolution pédagogique du système éducatif algérien. Comme les épreuves évaluant les compétences ne respectent pas les exigences de la nouvelle approche, elles ne deviennent pas valides pour une évaluation réelle des apprenants algériens. Une telle situation soulève de sérieuses interrogations quant aux raisons de la stagnation de l'évaluation scolaire : pour quelles raisons les outils évaluatifs ont-ils stagné ?

La première réponse qui s'impose à cette question -si nous tenons compte du lien étroit entre enseignement et évaluation- est la stagnation des pratiques enseignantes ; les pratiques évaluatives étant le reflet des pratiques enseignantes. Si les enseignants ne respectent pas les exigences de l'APC en évaluation du FLE, il devient évident que cette approche n'est pas



suffisamment intégrée dans leurs pratiques. Ammouden (2018), dans une étude menée auprès des enseignants algériens des trois cycles, affirme que la réforme est rejetée par ces enseignants qui considèrent que les nouveaux programmes et manuels ne répondent pas réellement aux besoins des apprenants vu les insuffisances constatées.

La deuxième raison qui explique la stagnation des épreuves évaluatives est la formation des enseignants qui est responsable également de la stagnation des pratiques enseignantes (Ammouden, 2019<sup>115</sup>). Certes, plusieurs séances de formation ont été programmées par les inspecteurs et les enseignants-formateurs pour que la nouvelle approche soit soigneusement appliquée par les enseignants. Cependant, ces séances de formation continue, axées sur les définitions des concepts et les pratiques enseignantes, ont carrément écarté les exigences de l'APC quant aux pratiques évaluatives, ce qui révèle une séparation surprenante de deux domaines indissociables : enseignement et évaluation.

La formation continue des enseignants ne doit pas ainsi négliger les pratiques évaluatives prônées par la nouvelle approche étant donné que le changement des pratiques évaluatives pourrait générer celui des pratiques enseignantes. Cette déclaration devient fondée si nous tenons compte des enseignants qui ont tendance à adapter leurs pratiques enseignantes aux démarches institutionnelles de l'évaluation pour garantir la réussite de leurs apprenants. Dans ce cas-ci, la réticence au changement et la fidélité à des démarches, techniques, stratégies et activités dépassées se trouvent récompensées dans les évaluations sommatives où les notes obtenues comptent plus que les compétences réelles. En préparant ces enseignants au changement des pratiques évaluatives, nous les poussons à changer leurs pratiques enseignantes pour que celles-ci soient conformes aux exigences. Si nous continuons à différer la réforme de l'évaluation, le changement n'arriverait jamais...

### **Conclusion**

En vue d'étudier la validité de l'évaluation sommative du FLE au cycle moyen, nous avons adopté la démarche analytique. Des épreuves sommatives (devoirs et compositions) de tous les niveaux, qui ont constitué notre premier corpus, ont été analysées pour vérifier si elles sont valides, autrement dit si elles évaluent réellement les compétences des collégiens algériens.

---

<sup>115</sup> L'auteur qui suite à une enquête menée auprès des enseignants algériens affirme que la formation des enseignants et celle des inspecteurs n'a pas atteint ses objectifs, d'où l'échec de la mise en œuvre de la nouvelle approche (APC).

Cette analyse s'est appuyée sur une grille comprenant plusieurs indicateurs relatifs aux outils de mesure et à ceux du jugement, inspirés de la littérature scientifique exploitée dans notre volet théorique.

Les résultats obtenus que nous avons exposés dans ce deuxième chapitre de notre volet pratique révèlent que les épreuves sommatives que les enseignants conçoivent pour évaluer les compétences de leurs apprenants en FLE ne sont pas valides étant donné que plusieurs exigences de l'APC et instructions officielles n'y sont pas respectées. Bien que cette approche préconise le recours à des situations inédites, complètes, motivantes...qui évaluent indissociablement compréhension et expression, les épreuves analysées continuent à les évaluer séparément, comme dans les épreuves traditionnelles, en recourant à des outils inadaptés aux nouvelles instructions : des items mesurent la compréhension (QCM, questions ouvertes...) et des situations mal conçues évaluent la production (ce sont des situations incomplètes, souvent travaillées en classe). Ces deux compétences ne sont pas les seules composantes évaluées étant donné que les outils grammaticaux –désignés par ressources dans notre étude- sont également mesurés à travers des items souvent décontextualisés. Cette évaluation des ressources dans une perspective sommative, qui ne doit théoriquement inclure que les ressources importantes, porte –dans les épreuves de notre corpus- sur des savoirs et savoir-faire. Ce sont ainsi les niveaux cognitifs inférieurs que les enseignants visent en évaluation, ce qui n'est pas conforme aux nouvelles instructions. Ces insuffisances qui affectent la validité ne sont pas exclusives aux outils de mesure. En effet, les outils de jugement -qui doivent être clairement mentionnés et communiqués aux apprenants- font défaut dans les épreuves analysées, ce qui conduit à la subjectivité des enseignants.

Les conclusions primaires auxquelles nous sommes parvenues suite à l'analyse menée reconnaissent que les outils de mesure et de jugement qui constituent les épreuves de FLE proposées pour évaluer les compétences des collégiens algériens ne sont pas valides étant donné que ces outils ne sont pas adaptés aux instructions de la nouvelle approche. Rappelons que la validité des épreuves nécessite celle des outils de mesure et de jugement intégrés. L'incapacité des pratiques évaluatives actuellement adoptées à refléter fidèlement les compétences des évalués s'explique par leur stagnation : ces pratiques-ci semblent en effet similaires à celles qui ont existé avant la réforme du système éducatif national.

En vue de vérifier cette stagnation des pratiques évaluatives, une deuxième démarche était adoptée : la confrontation des épreuves de BEM post-APC, censées être conformes aux

exigences, aux BEM pré-APC. Suite à cette comparaison, une similitude flagrante entre les deux catégories de BEM était démontrée : les outils de mesure y sont en effet quasi-identiques et les outils de jugement sont omis. La conclusion de stagnation des pratiques évaluatives, précédemment formulée, est ainsi prouvée.

Cette situation rend nécessaire une réforme urgente de l'évaluation scolaire pour concevoir des dispositifs évaluatifs valides donnant des aperçus fidèles sur les compétences des apprenants algériens en FLE, qui ne serait pas la seule discipline concernée par le changement étant donné que toutes les langues enseignées (arabe et anglais) sont évaluées suivant la même démarche. Cette réforme urgente nécessite une formation des enseignants, des formateurs et des inspecteurs pour les amener à intégrer pleinement les pratiques enseignantes préconisées par la nouvelle approche et les préparer à de nouvelles modalités d'évaluation.

# **CHAPITRE 3 : PROPOSITION DE DISPOSITIF ÉVALUATIF VALIDE**

## Introduction

Avant d'entamer notre dernier chapitre du volet pratique, nous rappelons brièvement les contenus des deux chapitres précédents. Le premier chapitre, où le cadre méthodologique (les approches adoptées, les outils de collecte investis, les corpus analysés...) est largement détaillé, précise que nous avons procédé à une analyse quantitative et qualitative des épreuves sommatives du FLE des dernières années scolaires (2017/2018, 2018/2019 et 2019/2020) pour déterminer à quel degré elles sont conformes aux exigences dictées par la nouvelle approche et aux instructions officielles diffusées suite à son application. Les résultats obtenus -présentés, analysés et discutés dans le deuxième chapitre pratique- ont révélé que les pratiques évaluatives sélectionnées dans les épreuves du cycle moyen ne respectent plus les exigences étant donné que les compétences (compréhension et expression) y sont évaluées séparément, que les outils de mesure intégrés ne fournissent pas des précisions sur les compétences effectives des évalués... ; ce qui signifie que les épreuves analysées ne sont pas valides pour évaluer les compétences des collégiens algériens.

Ces premiers résultats obtenus ont rendu nécessaire une confrontation des épreuves qui s'inscrivent dans une perspective de compétences à celles qui sont ancrées dans la pédagogie écartée (PPO) : les BEM pré-APC sont ainsi comparés aux BEM post-APC. Cette deuxième approche a montré que les épreuves sommatives ont stagné en dépit du changement méthodologique qu'a connu notre système éducatif, ce qui montre que les épreuves actuelles présentent des problèmes de validité. En effet, les outils de mesure et de jugement investis par les enseignants ne correspondent plus aux instructions officielles qui depuis la réforme du système éducatif en 2003 préconisent les situations d'évaluation comme instruments de mesure et les grilles critériées comme outils de jugement.

Vu ces résultats, il devient indispensable de recourir à des outils de mesure et de jugement qui tiennent compte des exigences de la nouvelle approche (APC). Pour les concevoir, nous nous référons à notre troisième chapitre théorique qui a listé les étapes à suivre pour la conception de dispositifs valides. Les références qui y sont citées (Roy, 2017...) énumèrent : la *planification*, la *conception*, la *passation*, le *jugement* et la *décision*. De toutes ces étapes, antérieurement exposées puis détaillées, nous retenons la *conception* et le *jugement* qui sont en lien avec l'élaboration des outils de mesure et de jugement. Ces étapes-ci qui sont particulières à la discipline à laquelle nous nous intéressons dans notre étude aboutiront à des situations évaluatives de FLE conformes aux exigences de l'APC et à des grilles critériées qui pourront évaluer convenablement les productions des apprenants.

La conception de chacun de ces deux outils suit des étapes antérieurement énumérées.

Pour les situations évaluatives, les phases retenues sont celles que propose Gérard (2009). Nous commencerons par déterminer les compétences à évaluer : toutes les compétences visées à travers les enseignements dispensés –à tous les niveaux du cycle- puis évaluées seront énumérées. Suite à cette précision, nous tenterons d'élaborer des situations évaluatives en tenant compte des instructions officielles : nous concevrons des contextes, sélectionnerons des supports à exploiter, proposerons des tâches à réaliser par les apprenants... Bien que des critères de correction doivent être déterminés -suite à la conception des situations-, nous les différerons à un autre moment de notre étude. En effet, nous les aborderons profondément lors de la conception des grilles critériées que nous proposerons pour la correction des productions obtenues. De ce fait, la conception des situations sera suivie d'une vérification de leur validité en s'appuyant sur la liste de vérification déjà présentée. Cette liste de vérification sera l'outil que nous soumettrons aux experts (inspecteurs du cycle moyen) que nous solliciterons au moment de l'expérimentation des situations conçues.

Le deuxième outil que nous tenterons de concevoir dans ce dernier chapitre est la grille de correction. Son élaboration suivra les étapes énumérées par Côté, Tardif et Munn (2011) : le choix des critères de correction, la sélection d'une échelle d'appréciation, la précision du moyen pour exprimer le jugement, l'assemblage de la grille et son expérimentation. Pour les critères de correction, qui doivent être pertinents, indépendants, pondérés et peu nombreux selon Gérard (2005, 2009), ils seront choisis en tenant compte de la particularité de la discipline évaluée et des instructions que contiennent les documents pédagogiques. Qu'ils soient des critères minimaux ou de perfectionnement, ils seront répartis en indicateurs qui fournissent des indications concrètes sur la maîtrise du critère. Suite à la précision des critères de correction retenus, nous sélectionnerons une échelle d'appréciation qui convient à notre visée sommative, ce qui permettra de déterminer le moyen à adopter pour exprimer le jugement. Lorsque la grille sera assemblée, nous la soumettrons à des experts qui en évaluent la pertinence et la clarté.

Mentionnons que les situations que contient ce chapitre ne sont que des propositions, qui pourraient être révisées, modifiées et enrichies par les enseignants qui les ajustent en fonction de leurs enseignements et de leurs objectifs.

## **1. Les compétences à évaluer au cycle moyen**

Pour identifier les compétences développées chez les apprenants du cycle moyen à travers les contenus sélectionnés puis évaluées à travers les outils précédemment décrits, nous nous référons au « *Document d'accompagnement du programme de français –Cycle moyen* » qui énumère les compétences (globale, terminales...) du cycle et propose des plans annuels, des situations d'intégration, des situations d'évaluation...

La compétence globale qui désigne ce que les apprenants devront maîtriser au bout du cycle (profil de sortie) y est ainsi annoncée : « *Au terme du cycle moyen, (...) l'élève est capable de comprendre et de produire à l'oral et à l'écrit des textes explicatifs, prescriptifs, descriptifs, narratifs, argumentatifs et ce, en adéquation avec la situation de communication* » (Commission nationale des programmes, 2016, p. 6). Pour que les apprenants soient capables de comprendre et de produire des énoncés oraux et écrits adaptés aux situations, les enseignants doivent les doter de connaissances, outils, stratégies... leur permettant de réussir les tâches prescrites dans les situations scolaires et les préparant pleinement à la complexité des situations authentiques.

Les types textuels retenus sont répartis sur les quatre niveaux du cycle moyen qui correspondent à trois paliers. Au niveau du premier palier (1AM), les types explicatif et prescriptif sont retenus. Pour le deuxième palier (2AM et 3AM), des textes narratifs –qui relèvent de divers genres- sont sélectionnés. Enfin, le dernier palier (4AM) propose des textes de divers types à visée argumentative. Pour chaque niveau, une compétence particulière est visée :

- Comprendre et produire *des textes explicatifs et prescriptifs* adaptés à la situation de communication est la compétence visée en 1AM.
- Comprendre et produire *des récits fictifs* (fables, légendes, contes...) en tenant compte des contraintes de la communication est la compétence visée en 2AM.
- Comprendre et produire *des récits de faits réels* en tenant compte de la situation de communication est la compétence visée en 3AM.
- Comprendre et produire *des textes à visée argumentative* en tenant compte des contraintes de la situation de communication est la compétence visée en 4AM.

Pour installer ces compétences, des ressources linguistiques sont enseignées aux apprenants –à travers des situations- pour être exploitées lors de la réalisation des tâches sélectionnées par les enseignants. Mentionnons que les situations d'apprentissage, d'intégration et d'évaluation investies traitent des thèmes précis (hygiène, progrès scientifique, découvertes, fléaux sociaux...).

## 2. Des propositions de situations évaluatives

Nombreux sont les documents officiels qui proposent des situations évaluatives supposées conformes aux exigences de la nouvelle approche (APC). Bien que ces situations éclairent les contextes et précisent les productions attendues, elles ne respectent pas toutes les exigences (absence des supports, nombre des tâches insuffisant...), d'où la nécessité de proposer des situations évaluatives plus adaptées aux instructions officielles. Nous avons pris l'initiative de produire quelques situations évaluant les compétences des collégiens algériens en FLE, des situations qui seront vérifiées et validées par des collègues exerçant au cycle moyen.

Pour chaque niveau, quatre situations sont produites pour évaluer une des compétences visées par les programmes officiels. Nous ne les exposerons pas toutes dans leur première version, à ce niveau de notre étude, et nous nous contentons de présenter une des situations conçues et analysées par les experts pour illustrer leurs commentaires, ci-dessous, exposés.

Voici une situation élaborée pour évaluer les apprenants de 4AM :

<b>Situation d'évaluation</b>	
<p>Pour célébrer la fête de l'Aïd-El-Adha, certaines familles organisent des barbecues dans les forêts, les montagnes ou les jardins. Malheureusement, certaines réunions familiales, qui s'achèvent dans la joie, aboutissent à des incendies involontaires qui ravagent parfois de vastes terrains. Cependant, ces incendies sont parfois volontairement causés par des ennemis de la nature pour collecter puis vendre le charbon de bois. Quelles qu'en soient les raisons, ces incendies deviennent de vrais problèmes qui menacent la vie des hommes et des animaux. Révolté par cette situation, tu décides-en collaboration avec tes amis du club <i>Amis de la nature</i>- d'élaborer des fiches à afficher sur les tableaux implantés dans les espaces verts les plus fréquentés de ta ville pour sensibiliser les visiteurs à la nécessité de les préserver. Comme tu t'occupes de la rédaction des textes à afficher, tu te réfères aux ressources suivantes pour :</p> <ul style="list-style-type: none"> <li>- Expliquer l'importance des arbres dans la vie des hommes.</li> <li>- Expliquer la nécessité de préserver des forêts dans la vie des animaux.</li> <li>- Proposer des conseils aux visiteurs qui permettent d'éviter les incendies des forêts.</li> </ul>	
<b>Document 1</b>	<b>Document 2</b>
<p>« <b>Les forêts abritent plus de la moitié de la biodiversité terrestre et l'on y découvre encore chaque jour de nouvelles espèces.</b> Elles sont un laboratoire où les hommes peuvent chercher de nouveaux médicaments. Elles fournissent du bois pour se chauffer. Elles abritent des animaux que l'on peut chasser pour se nourrir. Elles sont le lieu d'habitation de certaines tribus humaines. Elles sont utiles pour fabriquer des meubles, des maisons, du papier.</p>	<p>« <b>Les conséquences de la déforestation</b></p> <p><i>La perte de biodiversité</i> : Les forêts hébergent plus de 80 pour cent de la biodiversité terrestre et représentent l'un des derniers refuges pour de très nombreuses espèces animales et végétales. C'est pourquoi, la déforestation est une catastrophe aussi bien pour l'Homme que pour les autres espèces puisque on</p>



<p>Elles sont des espaces de repos, de promenade et de loisirs. Elles sont de véritables usines de dépollution de l'eau, du sol et de l'air. Elles absorbent du gaz carbonique et rejettent de l'oxygène. Par ce phénomène, elles participent à la lutte contre le réchauffement climatique.</p> <p><b>C'est assez difficile à imaginer, mais sans les forêts, la Terre deviendrait sans doute un désert.</b> Sans leur habitat, des centaines d'espèces d'animaux disparaîtraient. Il y aurait moins d'eau, moins d'oxygène. L'homme finirait sans doute lui-même par disparaître ».</p> <p>Site : <a href="https://www.pefc-france.org/123-foret/pourquoi-faut-il-protoger-les-forets/">https://www.pefc-france.org/123-foret/pourquoi-faut-il-protoger-les-forets/</a></p>	<p>estime que 27 000 espèces animales et végétales disparaissent chaque année à cause d'elle. Cette perte de biodiversité, qui peut être irréversible, coupe l'humanité de services et ressources inestimables. En effet, les systèmes alimentaires sont fortement dépendants de la biodiversité et une proportion considérable de médicaments est directement ou indirectement d'origine biologique ».</p> <p>Site : <a href="https://www.notre-planete.info/environnement/deforestation.php">https://www.notre-planete.info/environnement/deforestation.php</a></p>
<p style="text-align: center;"><b>Document 3</b></p> <p>« Le déboisement a donc des conséquences dramatiques pour les humains et leur environnement : il libère du carbone emmagasiné par les forêts tout au long de leur vie. Le réchauffement climatique affecte à son tour les écosystèmes en limitant leur capacité à stocker davantage de carbone. Les forêts jouent aussi un rôle clé dans l'équilibre fragile du cycle de l'eau.</p> <p>On comprend ainsi que <b>les forêts sont indispensables à l'équilibre climatique</b> et qu'il faut à tout prix les préserver.</p> <p>Si les forêts doivent être protégées, c'est aussi pour <b>préserver la biodiversité irremplaçable d'espèces animales et végétales qu'elles abritent</b>, ainsi que les Hommes qui en dépendent. Un tiers de la population mondiale utilise aujourd'hui du bois et des ressources naturelles issues de la forêt pour ses <b>besoins vitaux (construction, cuisine, chauffage, activités génératrices de revenus notamment)</b>. La déforestation menace donc la vie de communautés entières, leurs territoires ancestraux, la nature et les services écosystémiques dont elles dépendent et les savoirs traditionnels dont elles sont détentrices ».</p> <p>Site : <a href="https://planete-urgence.org/agir-foret-biodiversite/">https://planete-urgence.org/agir-foret-biodiversite/</a></p>	<p style="text-align: center;"><b>Document 4</b></p>  <p>« Les forêts :</p> <ul style="list-style-type: none"> <li>▪ captent l'énergie solaire,</li> <li>▪ séquestrent le carbone,</li> <li>▪ dégagent de l'oxygène,</li> <li>▪ font évaporer l'eau,</li> <li>▪ filtrent l'eau du sol,</li> <li>▪ enrichissent et préservent les sols,</li> <li>▪ guérissent et alimentent,</li> <li>▪ abritent diverses espèces.</li> </ul> <p>Avec tous les services et le bien que les forêts nous apportent, il est primordial de les protéger et de les conserver. La préservation des forêts apparaît comme un enjeu essentiel ».</p> <p>Site : <a href="https://news.all4trees.org/importance-preserver-forets/">https://news.all4trees.org/importance-preserver-forets/</a></p>

### 3. Vérification des situations conçues

Pour vérifier si les situations évaluatives, ci-dessus exposées, correspondent réellement aux exigences de la nouvelle approche (APC), nous les avons confrontées aux indicateurs de la liste de vérification élaborée qui énumère les principales caractéristiques de ces outils de mesure.

- Les situations sont-elles complexes ?

Toutes les situations, qui mobilisent les ressources antérieurement enseignées, présentent des problèmes que les apprenants doivent résoudre en s'appuyant sur les supports

textuels proposés. Il est clair que ces outils qui nécessitent une analyse de ces supports mobilisent certaines opérations cognitives de haut niveau. Les apprenants sont en effet contraints d'analyser les informations fournies, les évaluer pour repérer les plus utiles puis les investir dans de nouvelles productions.

- Sont-elles signifiantes ?

Elles présentent des problèmes issus du vécu des apprenants les poussant ainsi à s'engager pour les résoudre. Au niveau de ce critère, notre jugement reste subjectif étant donné que ce qui nous semble significatif pourrait ne pas l'être pour une autre personne. Il en est de même pour les apprenants étant donné que ce qui est significatif pour un apprenant ne l'est pas pour un autre.

- Sont-elles authentiques ?

Ces situations présentent des contextes vraisemblables que les apprenants pourraient confronter dans leur vie quotidienne. Les tâches réalistes qu'elles proposent pourraient nous donner un aperçu sur les performances de ces apprenants évalués dans les situations sociales et réelles. Cependant, cet aperçu demeure approximatif étant donné que de nombreux paramètres entrent en jeu et déterminent la performance d'un apprenant donné dans une situation donnée.

- Sont-elles complètes ?

De toutes les caractéristiques, celle-ci est la plus pertinente et la plus aisée à vérifier. En effet, elle est déclinée en éléments clairement observables dans les situations conçues. Celles-ci mentionnent les productions évaluables attendues des apprenants et leurs buts. Les contextes y sont également explicités pour que les apprenants puissent élaborer des productions pertinentes. Pour les supports, leurs thèmes correspondent à ceux que les programmes proposent. Cependant, leur nombre est suggéré étant donné que les situations évaluatives que les manuels et les enseignants proposent les excluent. Mentionnons que ce nombre correspond à celui des supports que contient l'épreuve certificative- citée dans notre étude- extraite du « *Document d'accompagnement du programme de français –Cycle* ». Quant aux consignes, elles sont claires et précises annonçant des tâches qui aboutissent à des productions concrètes. Pour qu'elles soient conformes aux exigences, trois tâches indépendantes sont sélectionnées dans chacune des situations proposées.

Bien que les situations contiennent les composantes précédentes, elles omettent trois éléments : les conditions de réalisation, les instructions de réalisation (la longueur, les ressources à mobiliser...) et les grilles de correction. Si le premier élément dépend des choix des enseignants et des contraintes institutionnelles, les deux derniers sont à insérer pour que les situations soient complètes et valides.

- Sont-elles inédites ?

Ce sont des situations distinctes de celles que les manuels proposent. D'ailleurs, elles sélectionnent des supports qui ne figurent pas dans les manuels scolaires actuels.

- Sont-elles adéquates ?

Les tâches à résoudre que les situations proposent évaluent les compétences que les programmes du cycle moyen visent. La longueur des supports, la clarté du vocabulaire... sont à adapter à la particularité des apprenants à évaluer.

De ce qui précède, nous pouvons affirmer que les situations conçues respectent de nombreuses exigences de l'évaluation des compétences. Cette vérification, que nous avons effectuée nous-mêmes, est suivie par une expérimentation effectuée par d'autres enseignants exerçant au cycle et ce, en s'appuyant sur la même grille.

#### **4. Expérimentation**

En vue de vérifier si les situations que nous avons conçues sont conformes aux exigences précédemment mentionnées, nous recourons à la consultation des experts<sup>116</sup>, une des démarches retenues –par des auteurs comme Gérard et De Ketele (2005), Gérard (2009)...- pour expérimenter les outils de mesure. Nous leur soumettons la liste de vérification qui explicite toutes les caractéristiques que les situations évaluatives doivent remplir. Il est important de mentionner à ce niveau que le prétest -qui consiste à expérimenter les situations évaluatives auprès des apprenants- est évité vu que les pratiques évaluatives actuelles ne correspondent pas à celles que préconisent la nouvelle approche et les instructions officielles.

Suite à cette expérimentation, les commentaires suivants sont émis :

- Préciser les compétences à évaluer : les situations soumises aux collègues ne déterminent pas les compétences évaluées ; d'où la nécessité de les préciser pour que les enseignants puissent les évaluer correctement.
- Ajouter certaines précisions nécessaires : constatant que les situations proposées ne contiennent que les supports et les textes explicitant les contextes, les consignes et les produits attendus, les experts consultés exigent l'ajout de certaines précisions que toute épreuve doit contenir. Il s'agit du type de l'épreuve et de sa durée, du trimestre, du niveau des apprenants concernés...

---

<sup>116</sup>Il convient de préciser que les experts consultés pour la vérification des situations produites sont des inspecteurs de l'éducation nationale. Leurs rapports sont insérés dans la partie Annexes.

- Rappeler les ressources linguistiques à investir : bien que cet élément ne soit pas indispensable, étant donné que certains enseignants désirent vérifier si leurs apprenants peuvent déterminer les ressources appropriées pour résoudre les situations proposées ; les experts consultés tenant compte du niveau des évalués jugent préférable de les mentionner pour que les apprenants les investissent dans les productions à réaliser.
- Insérer les instructions de réalisation : omises également, ces instructions fournissent des informations techniques sur les textes à produire telles que leur structure, leur longueur...
- Préciser les modalités de réponse : comme les situations conçues ne précisent pas les modalités de réponses, il est important de les ajouter notamment que des espaces ne sont pas prévues pour que les apprenants insèrent leurs productions.
- Insérer des notes pour expliquer ce qui est difficile : en sélectionnant les ressources à fournir aux apprenants, les enseignants doivent veiller à une caractéristique essentielle mentionnée dans notre liste de vérification : adaptation des supports au niveau des apprenants. Suite à la vérification de quelques supports, les experts consultés notent la complexité du vocabulaire investi par les auteurs et suggèrent l'insertion de notes explicatives qui définissent les mots jugés difficiles, une suggestion adoptée dans les situations proposées.
- Veiller sur l'indépendance des consignes : il est connu que les tâches sélectionnées doivent être indépendantes pour la validité des épreuves conçues. Cette exigence fait défaut, selon les experts, dans certaines consignes ; d'où leur reformulation dans la version finale proposée.


Ces commentaires sont pris en considération et des modifications sont apportées aux situations évaluées par les experts consultés<sup>117</sup>. Nous les exposerons dans le dernier élément de ce chapitre pratique étant donné que celles-ci sont à compléter par une grille de correction explicitant les critères à adopter en FLE. Seule la version finale de la situation-exemple exposée au début de notre travail sera présentée pour illustrer les changements apportés suite aux remarques des experts.

---

<sup>117</sup>Seuls les commentaires fondés ont été pris en considération. Certains commentaires qui suggèrent le recours aux situations conçues comme des situations d'apprentissage et d'intégration et de maintenir les épreuves sommatives actuelles n'étaient pas pris en considération étant donné que toutes les situations doivent appartenir à une même famille ; ce qui suppose des paramètres identiques.

<b>Compétence évaluée</b>	
Comprendre et produire <b>des textes argumentatifs</b> en tenant compte des contraintes de la situation de communication	
<b>Épreuve<sup>118</sup></b>	
<b>Établissement</b> : ..... <b>Type</b> : ..... <b>Matière</b> : ..... <b>Niveau</b> : ..... <b>Durée</b> : .....	
<b>Évaluation</b>	
<p>Pour célébrer la fête de l'Aïd-El-Adha, certaines familles algériennes organisent des barbecues dans les forêts, les montagnes ou les jardins. Malheureusement, certaines réunions familiales, qui s'achèvent dans la joie, aboutissent à des incendies involontaires qui ravagent parfois de vastes terrains. Cependant, ces incendies sont parfois volontairement causés pour collecter puis vendre du charbon de bois. Quelles qu'en soient les raisons, ces incendies deviennent de vrais problèmes qui menacent la vie des hommes et des animaux. Révolté par cette situation, tu décides d'élaborer des affiches à exposer dans les espaces verts les plus fréquentés de ta ville pour sensibiliser les visiteurs à la nécessité de les préserver. En te référant aux documents ci-dessous proposés, écris trois textes de 10 lignes pour :</p> <ul style="list-style-type: none"> <li>- Expliquer l'importance des arbres dans la vie des hommes.</li> <li>- Expliquer la nécessité de préserver les forêts dans la vie des animaux.</li> <li>- Proposer des conseils aux visiteurs pour éviter les incendies des forêts.</li> </ul>	
<b>Document 1</b>	<b>Document 2</b>
<p>« <b>Les forêts abritent plus de la moitié de la biodiversité terrestre et l'on y découvre encore chaque jour de nouvelles espèces.</b> Elles sont un laboratoire où les hommes peuvent chercher de nouveaux médicaments. Elles fournissent du bois pour se chauffer. Elles abritent des animaux que l'on peut chasser pour se nourrir. Elles sont le lieu d'habitation de certaines tribus humaines. Elles sont utiles pour fabriquer des meubles, des maisons, du papier. Elles sont des espaces de repos, de promenade et de loisirs. Elles sont de véritables usines de dépollution de l'eau, du sol et de l'air. Elles absorbent du gaz carbonique et rejettent de l'oxygène. Par ce</p>	<p>« <b>Les conséquences de la déforestation</b></p> <p><i>La perte de biodiversité</i> : Les forêts hébergent plus de 80 pour cent de la biodiversité terrestre et représentent l'un des derniers refuges pour de très nombreuses espèces animales et végétales. C'est pourquoi, la déforestation est une catastrophe aussi bien pour l'Homme que pour les autres espèces puisque on estime que 27 000 espèces animales et végétales disparaissent chaque année à cause d'elle. Cette perte de biodiversité, qui peut être irréversible, coupe l'humanité de services et ressources inestimables. En effet, les systèmes alimentaires sont</p>

<sup>118</sup> Dans cette situation rectifiée, certaines précisions sur l'épreuve (type, durée, niveau...), la compétence évaluée, la longueur des textes à produire, les ressources linguistiques à investir... sont ajoutées. Concernant les critères de correction, nous préférons ne pas les détailler car nous proposerons une grille de correction très précise.

<p>phénomène, elles participent à la lutte contre le réchauffement climatique.</p> <p><b>C'est assez difficile à imaginer, mais sans les forêts, la Terre deviendrait sans doute un désert.</b> Sans leur habitat, des centaines d'espèces d'animaux disparaîtraient. Il y aurait moins d'eau, moins d'oxygène. L'homme finirait sans doute lui-même par disparaître ».</p> <p><i>PEFC, « Pourquoi faut-il protéger les forêts ? » Dans <a href="http://www.pefc-france.org">www.pefc-france.org</a>, Consulté en 2019.</i></p>	<p>fortement dépendants de la biodiversité et une proportion considérable de médicaments est directement ou indirectement d'origine biologique ».</p> <p><i>Christophe Magdelaine, « La déforestation : définition, données, causes et conséquences », Dans <a href="http://www.notre-planete.info">www.notre-planete.info</a>, publié le 11/12/2020.</i></p>
<p style="text-align: center;"><b>Document 3</b></p> <p>« Le déboisement a donc des conséquences dramatiques pour les humains et leur environnement : il libère du carbone emmagasiné par les forêts tout au long de leur vie. Le réchauffement climatique affecte à son tour les écosystèmes en limitant leur capacité à stocker davantage de carbone. Les forêts jouent aussi un rôle clé dans l'équilibre fragile du cycle de l'eau.</p> <p>On comprend ainsi que <b>les forêts sont indispensables à l'équilibre climatique</b> et qu'il faut à tout prix les préserver.</p> <p>Si les forêts doivent être protégées, c'est aussi pour <b>préserver la biodiversité irremplaçable d'espèces animales et végétales qu'elles abritent</b>, ainsi que les Hommes qui en dépendent. Un tiers de la population mondiale utilise aujourd'hui du bois et des ressources naturelles issues de la forêt pour ses <b>besoins vitaux (construction, cuisine, chauffage, activités génératrices de revenus notamment)</b>. La déforestation menace donc la vie de communautés entières, leurs territoires ancestraux, la nature et les services écosystémiques dont elles dépendent et les savoirs traditionnels dont elles sont détentrices ».</p> <p><i>Association Le sens du partage, Appel pour lutter contre le réchauffement climatique, Dans <a href="https://lesensdupartage.org">https://lesensdupartage.org</a>, 2008.</i></p>	<p style="text-align: center;"><b>Document 4</b></p>  <p>« Les forêts :</p> <ul style="list-style-type: none"> <li>▪ captent l'énergie solaire,</li> <li>▪ séquestrent le carbone,</li> <li>▪ dégagent de l'oxygène,</li> <li>▪ font évaporer l'eau,</li> <li>▪ filtrent l'eau du sol,</li> <li>▪ enrichissent et préservent les sols,</li> <li>▪ guérissent et alimentent,</li> <li>▪ abritent diverses espèces.</li> </ul> <p>Avec tous les services et le bien que les forêts nous apportent, il est primordial de les protéger et de les conserver. La préservation des forêts apparaît comme un enjeu essentiel ».</p> <p><i>Nathalie Tachet, « Les forêts, pourquoi il est important de les préserver ? » Dans <a href="https://news.all4trees.org">https://news.all4trees.org</a>, publié en 2019.</i></p>
<p>Dans tes productions, il est important de mettre les verbes au présent de l'indicatif, employer les connecteurs logiques et chronologiques, utiliser une ponctuation appropriée et respecter la structure des textes à visée argumentative.</p>	
<p><b>Les critères de correction</b>, que contient la grille de correction, sont à communiquer aux apprenants.</p>	

## 5. Proposition de grilles de correction

Pour concevoir les grilles de correction, nous nous référons aux étapes énumérées par Côté, Tardif et Munn (2011) et listées dans notre volet théorique. La première étape que nous précisons en premier lieu est la sélection des critères de correction.

- Les critères retenus en FLE

Les critères retenus pour évaluer le FLE –qui rappelons-le ressemblent à ceux retenus pour évaluer les langues (Roegiers, 2005)- sont identiques aux critères basiques applicables à toutes les disciplines : la **pertinence**, la **correction** et la **cohérence**. Ces critères minimaux que listent les documents d’accompagnement des quatre niveaux du cycle se réfèrent à des travaux menés par Roegiers (2004), Gérard (2005, 2009)... Quant aux critères de perfectionnement que ces documents officiels ne proposent pas, nous retenons **l’originalité** de la production.

Chacun de ces critères est défini par des indicateurs qui varient selon les tâches prescrites et les productions attendues. Cependant, certains indicateurs généraux pourraient être retenus. Ainsi, le respect de la consigne et l’utilisation des ressources sont des éléments qui définissent la pertinence de la production (indicateurs retenus dans notre grille). Gérard (2009), qui définit la pertinence comme une interprétation correcte de la situation, propose le respect de la consigne ; la réponse correcte au contexte ; la réponse à ce qui est attendu et le choix des idées adéquates. Tous ces éléments cités par l’auteur nous semblent étroitement liés au respect de la consigne, un indicateur déjà retenu.

Le critère de la correction semble peu détaillé dans les documents officiels consultés. En effet, il est décrit par un indicateur unique qui consiste au respect de la syntaxe et des normes morpho-grammaticales, un indicateur qui ne nous semble pas assez clair, ni directement observable dans les productions à corriger. Cette clarté, nous la retrouvons dans les indicateurs proposés par Gérard (2009) qui estime que l’utilisation correcte des outils linguistiques, autrement dit la correction de la langue, pourrait être observée à travers : le respect des règles, la construction correcte des phrases, l’utilisation des signes de ponctuation, la clarté du vocabulaire... D’autres indicateurs qui explicitent ce critère sont également proposés par Ruzibiza et De Ketele (2007) dans les grilles que leur article englobe. Il s’agit de : la structuration syntaxique des phrases, la correction orthographique des mots, la clarté du vocabulaire et les connecteurs logiques. En nous inspirant de ces auteurs, nous retenons : la structuration syntaxique des phrases, la correction orthographique, la correction de la conjugaison et la clarté du vocabulaire ; qui correspondent aux ressources souvent enseignées aux apprenants.

Le dernier critère minimal, la cohérence, semble le plus explicité dans les documents officiels. Les indicateurs qui le décrivent- en dépit de leur légère variété d’un niveau à un autre- sont souvent : le respect de la structure des types textuels à rédiger, utilisation des temps verbaux (deux indicateurs retenus dans notre grille), utilisation des ressources linguistiques... Il est clair que certains indicateurs sont introduits dans des critères différents tels l’utilisation des

ressources retenu au niveau de la pertinence et de la cohérence ; d'où la nécessité de sélectionner minutieusement les indicateurs relevant de chacun des critères sélectionnés. Mentionnons à ce niveau que les enseignants pourraient intégrer des éléments observables (indicateurs) qui ne figurent pas dans les documents institutionnels tels que la continuité des idées exprimées, la non-contradiction...

Ces critères retenus dans la grille dans la grille proposée sont des critères minimaux dont la maîtrise est nécessaire pour que la compétence soit maîtrisée. Nous les complétons par un dernier critère de perfectionnement : l'originalité. Ce dernier critère semble complètement négligé dans les documents officiels consultés qui ignorent les critères de perfectionnement et se contentent de mentionner les critères minimaux et les indicateurs qui en découlent. Nous retenons néanmoins deux indicateurs : la nouveauté de la production et son caractère personnel.

- L'échelle d'appréciation et ses échelons

Dans notre volet théorique, de nombreuses catégories d'échelles sont exposées. De toutes ces catégories, les échelles numériques qui constituent une sous-catégorie des échelles uniformes quantitatives sont retenues dans la mesure où elles aboutissent à des notes, les moyens les plus appropriés et les plus courants pour exprimer le jugement dans notre système éducatif. Ces échelles numériques seront constituées d'échelons représentant le niveau de maîtrise de chacun des critères sélectionnés par les enseignants.

Dans la grille que nous proposerons, nous retenons trois échelons (0-1-2) correspondant à une maîtrise insuffisante, maîtrise suffisante et maîtrise maximale.

- Pour exprimer le jugement...

Des deux outils exprimant le jugement, les notes constituent les outils les plus investis et notamment les plus pratiques, comme l'affirme De Vecchi (2011). Comme le préconise Gérard (2009), nous appuyons la note à laquelle aboutit la correction par des expressions phrastiques qui précisent si les critères minimaux sont maîtrisés ; ce qui est indispensable à la maîtrise de la compétence.

- La grille obtenue...

Après la sélection des critères, des indicateurs, de l'échelle d'appréciation et des outils exprimant le jugement, nous parvenons à une phase très importante : l'assemblage de la grille, que nous effectuerons en nous référant aux précautions listées par Côté, Tardif et Munn (2011) concernant la présentation et l'organisation des grilles de correction.

<b>Grille de correction</b>
<b>Épreuve</b>



Type : .....											
Matière : .....											
Date : .....											
<b>Correcteur</b>											
Nom : .....											
Prénom : .....											
Établissement : .....											
<b>Apprenant</b>											
Nom : .....											
Prénom : .....											
Groupe : .....											
<b>Évaluation des productions</b>											
Critères	Indicateurs	Tâche 1			Tâche 2			Tâche 3			Total
		0	1	2	0	1	2	0	1	2	
<b>Pertinence</b>	Respect de la consigne										
	Utilisation des ressources										
<b>Correction</b>	Structuration syntaxique des phrases										
	Correction orthographique										
	Clarté du vocabulaire										
	Correction de la conjugaison										
<b>Cohérence</b>	Respect de la structure des textes à rédiger										
	Utilisation des temps verbaux										
<b>Originalité</b>	Nouveauté des productions										
	Idées personnelles										
<b>Note obtenue</b>					..... /20						
<b>Seuil de réussite</b>					Le critère 1 est maîtrisé dans 2 tâches Le critère 2 est maîtrisé dans 2 tâches Le critère 3 est maîtrisé dans 2 tâches La note obtenue dépasse la moyenne (10)						
<b>Jugement</b>					<b>Réussite</b>			<b>Echec</b>			
*La bonne réponse est à cocher											

Comme cette grille est nouvellement introduite, des précisions s'imposent pour que les enseignants puissent l'exploiter convenablement.

La maîtrise des critères minimaux dans les trois tâches est évaluée séparément. Chaque critère est maîtrisé si les indicateurs pourraient être observés dans les productions des apprenants. Lorsque tous les indicateurs y sont présents, la maîtrise du critère est maximale et la note 02 est attribuée. Quand la moitié des indicateurs y sont présents, la maîtrise du critère est suffisante et la note 01 est attribuée. Dans le cas où les indicateurs présents sont rares ou

inexistants, la note 00 est accordée. Si la maîtrise du critère est suffisante dans deux tâches au moins, le critère est considéré maîtrisé. Contrairement aux critères minimaux, le seul critère de perfectionnement -évalué une seule fois- est noté selon le niveau de sa maîtrise, tout comme les critères minimaux.

La note attribuée par l'apprenant ne constitue pas la seule condition pour sa réussite étant donné que la maîtrise des critères minimaux est indispensable à celle de la compétence évaluée.

- Expérimentation de la grille proposée

Pour la validation de la grille conçue, nous recourons à la consultation des collègues, la procédure empruntée pour valider les situations. Cette validation s'appuie sur la liste de vérification qui intègre des indicateurs relatifs aux grilles de correction. Contrairement aux situations, où les commentaires des experts étaient nombreux, la grille proposée –vu sa similitude aux grilles proposées dans les références théoriques citées et dans les documents officiels- est validée par les collègues consultés.

## 6. Dispositif évaluatif proposé

Cet élément expose les évaluations sommatives que nous proposons pour évaluer les compétences des collégiens algériens en FLE. Ce sont des situations qui pourraient être ajustées en tenant compte des objectifs institutionnels, des publics évalués, de leurs intérêts...

### 6.1. Première année moyenne


#### Situation 01

<b>Compétence évaluée</b>
Comprendre et produire des textes explicatifs en tenant compte des contraintes de la situation de communication <sup>119</sup>
<b>Épreuve</b>
<b>Établissement</b> : .....
<b>Type</b> : .....
<b>Matière</b> : .....
<b>Niveau</b> : .....
<b>Durée</b> : .....
<b>Évaluation</b>

<sup>119</sup> Les énoncés de compétences visées puis évaluées au cycle moyen sont extraits du document d'accompagnement (2016). Contrairement aux énoncés de ce document où les compétences de compréhension et de production sont dissociées (conformément à la démarche précédente), nous les avons regroupées dans nos énoncés.

Ta mère t'a demandé de l'accompagner pour faire les courses. Elle t'a chargé(e) d'apporter du savon liquide, du dentifrice et des cotons-tiges. Tu ne lui apportes cependant que les deux premiers et tu lui expliques que les cotons-tiges ne doivent pas être utilisés car ce n'est pas un bon moyen pour nettoyer les oreilles. Ta mère ainsi que la vendeuse –intéressée par tes propos- semblent peu convaincues. Quand tu rentres à la maison, tu écris sur ta page Facebook : « Ma mère a de nouveau acheté des cotons-tiges ! ». Trois de tes amis réagissent par des questions intéressantes. Réponds-leur par des textes de 8 lignes pour expliquer :

- Pourquoi il faut éviter l'utilisation des cotons-tiges.
- Comment on peut nettoyer ses oreilles.
- Pourquoi le nettoyage exagéré des oreilles est déconseillé.

Document 1	Document 2
<p>« L'eau savonneuse est la solution la plus simple et la moins coûteuse pour nettoyer le <b>conduit auditif</b> sans risque. Lorsque l'on prend sa douche ou son bain, avec le jet de la <b>pomme de douche</b>, il suffit d'introduire l'eau dans l'oreille. Il faut néanmoins faire très attention à la puissance du jet (pour ne pas abîmer le tympan) et à la température de l'eau (qui doit être à 37°C, comme le corps humain) afin de ne pas provoquer de vertige.</p> <p>On peut également utiliser une <b>poire</b> en remplacement de la pomme de douche. Après avoir bien évacué l'eau, essuyez délicatement l'entrée du conduit auditif avec une serviette de toilette afin de sécher l'oreille ».</p>  <p style="text-align: center;"><i>Poire de nettoyage de l'oreille</i></p> <p><i>Comment bien se nettoyer les oreilles ? Dans <a href="https://audition44.fr">https://audition44.fr</a>, Consulté en 2020.</i></p> <p>*<b>Conduit auditif</b> : canal à travers lequel les sons passent.</p> <p>*<b>Pomme de douche</b> : élément qui assure une bonne diffusion de l'eau de la douche.</p> <p>*<b>Poire</b> : accessoire à remplir de liquide pour nettoyer les oreilles.</p>	<p>« Le mauvais usage des <b>cotons-tiges</b> a entraîné des saignements (35 %) et des douleurs aux oreilles (17 %). Dans un tiers des cas, un corps étranger était détecté, une déchirure du tympan pouvant mener à une perte d'audition était relevé un quart des cas. Dans plus de trois quarts des accidents, les enfants étaient eux-mêmes à l'origine de leur blessure. Le nettoyage de l'oreille était la cause la plus fréquente des visites aux urgences. Ce n'est pas le coton-tige qui est dangereux, mais sa mauvaise utilisation. « <i>Normalement, il faut utiliser le coton-tige seulement à l'extérieur de l'oreille, dans la partie visible par une autre personne</i> », souligne le professeur Dominique Chevalier, secrétaire général de la Société française d'ORL, au Figaro. Or le coton-tige est très souvent inséré trop profondément dans l'oreille ».</p> <p><i>Candice Satara-Bartko, Coton-tige : trop de blessures chez les enfants, <a href="https://paroledeamans.com">https://paroledeamans.com</a>, publié en 2017.</i></p> <p>*<b>Coton-tige</b> : bâtonnet entouré de coton pour nettoyer les oreilles.</p>
<p style="text-align: center;"><b>Document 3</b></p> <p><b>Se laver les oreilles une fois par semaine maximum</b></p> <p>« Pour une bonne santé des oreilles, évitez de vous laver les oreilles trop souvent. Comme précisé ci-dessus, vous risqueriez de priver l'oreille de <b>cérumen</b>, sa meilleure protection. Une fois par semaine - voire une fois tous les dix jours - est un rythme raisonnable et recommandé par les experts. En cas d'<b>écoulements</b>, vous pouvez passer un mouchoir humide sur le <b>pavillon</b> ».</p>	<p style="text-align: center;"><b>Document 4</b></p> <p>« Une bonne hygiène des oreilles consiste à ne pas agresser l'oreille, que ce soit avec des produits d'hygiène ou des accessoires inadaptés ».</p> <p><i>Comment bien se laver les oreilles ? Dans <a href="https://ideal-audition.fr">https://ideal-audition.fr</a>, Consulté en 2020</i></p>

<p><i>Maylis Choné, Comment bien se laver les oreilles ? Dans <a href="https://passeportsante.net">https://passeportsante.net</a>, Consulté en 2020.</i></p> <p>*<b>Cérumen</b> : matière jaune formée à l'intérieur de l'oreille.</p> <p>*<b>Écoulements</b> : sortie de liquides de l'oreille.</p> <p>*<b>Pavillon</b> : la partie extérieure et visible de l'oreille.</p>	
<p>Pour rédiger tes textes : conjugue les verbes au présent de l'indicatif, emploie les connecteurs d'énumération et respecte la structure du texte explicatif.</p>	
<p><b>Les critères de correction</b>, que contient la grille de correction, sont à communiquer aux apprenants.</p>	

### Situation 02

Compétence évaluée	
Comprendre et produire des textes explicatifs en tenant compte des contraintes de la situation de communication	
Épreuve	
Établissement : .....	
Type : .....	
Matière : .....	
Niveau : .....	
Durée : .....	
Évaluation	
<p>Le goûter de ton petit frère, qui est un écolier, est constitué souvent de gâteaux, chocolats et de jus. Ces aliments ont déséquilibré son régime alimentaire : il a supprimé le repas du déjeuner et il a réduit les aliments du dîner. Craignant les maladies nutritionnelles, tes parents ont consulté un spécialiste qui lui a prescrit un régime alimentaire où d'autres aliments remplacent les gâteaux. Constant que le cas de ton frère n'est pas le seul, tu décides de réaliser des brochures destinées aux élèves de son école pour les sensibiliser à la nécessité de bien manger. Tu commenceras ton projet par rédiger trois textes, de huit (08) lignes chacun, qui précisent :</p> <ul style="list-style-type: none"> <li>- Pourquoi il faut éviter un goûter constitué de gâteaux, de chocolats et de jus.</li> <li>- Les avantages d'une alimentation variée.</li> <li>- Les aliments constituant un goûter sain, en proposant un exemple.</li> </ul> <p>En vue de répondre à ces questions, tu exploites les documents ci-dessous qui fournissent quelques informations sur le thème traité.</p>	
<p style="text-align: center;"><b>Document 1</b></p> <p>« Le goûter est souvent le repas préféré des enfants. Mais que leur donner à manger pour en faire une pause à la fois gourmande et équilibrée ?</p>	<p style="text-align: center;"><b>Document 2</b></p>



La composition d'un goûter équilibré chez un enfant varie selon son âge, ses activités et de son appétit.

### De quoi est fait un goûter équilibré ?

Un goûter équilibré se compose d'un produit céréalier comme le pain, les biscuits, les céréales, d'un produit laitier au choix : verre de lait, portion de fromag, yaourt...et d'un fruit », explique Catherine Bourron-Normand, diététicienne. La quantité de pain, de biscuits, de fruit diffère selon les âges.

*Véronique Bertrand & Laura Annaert, Pour un goûter enfant équilibré, que faut-il manger ? Dans [www.santemagazine.fr](http://www.santemagazine.fr), publié le 5/2/2017*

\***Goûter** : repas léger pris en fin d'après-midi, ou entre le petit-déjeuner et le déjeuner.



« Manger de façon variée est l'une des clés pour limiter l'apparition de maladies. Ainsi, votre corps aura tous les nutriments indispensables pour son bon fonctionnement. Les aliments sains notamment les fruits et légumes vont permettre de **renforcer le système immunitaire** et **garder de l'énergie**. Toute l'importance d'une alimentation variée se ressentira également au niveau de votre forme physique. Un corps rempli des bons aliments est un corps à même de fournir tout ce qu'il faut pour **rester en forme**.

Dernier conseil à retenir : en complément d'une bonne alimentation diversifiée comme nous vous l'expliquions précédemment, pensez à **entretenir votre forme physique en faisant du sport**. Marche, course à pied, vélo, musculation, sports collectifs... Tous les moyens sont bons pour améliorer votre forme physique et aider votre corps à fonctionner de la meilleure des manières possibles et imaginables ». *Pourquoi il est important de manger varié ? Dans [www.calculersonimc.fr](http://www.calculersonimc.fr), Consulté en 2020*

### Document 3

#### « Bon à savoir

Gâteaux, biscuits, barres chocolatées, bonbons, sodas... dès le plus jeune âge, les enfants consomment beaucoup trop de sucre : «75 % des 4-7 ans, 60 % des 8-12 ans et 25 % des 13 - 17 ans en consomment trop», alerte l'Anses. Or, des apports excessifs présentent des risques pour la santé. À long terme, ils peuvent favoriser certaines maladies comme les caries dentaires, le diabète, l'obésité, ou la stéatose hépatique non-alcoolique (une inflammation du foie liée à une consommation excessive de sucres, d'où son nom de *maladie du foie gras* ou de *maladie du soda*) ».

*7 conseils pour limiter la consommation de sucre de vos enfants, Dans <https://assurance.carrefour.fr>, publié le 6/10/2020*

### Document 4

#### « Comment bien manger ?

La grande majorité des spécialistes en alimentation, qu'ils soient ou non du milieu officiel, s'entendent sur un certain nombre de principes qui peuvent nous servir de guides, les voici :

Une alimentation équilibrée : il est conseillé de choisir ses aliments dans les différents groupes alimentaires : les légumes et les fruits (la moitié de l'assiette), les produits céréaliers (le quart de l'assiette), les viandes et les substituts (l'autre quart), auxquels on ajoutera un apport de calcium, en consommant des produits laitiers, par exemple. Ainsi, votre repas contiendra : une bonne dose de glucides, suffisamment de protéines et peu de lipides.

Une alimentation variée : pour atteindre l'éventail nécessaire de nutriments et éviter les carences, il faut non seulement consommer chaque jour des aliments de chaque groupe alimentaire, mais plusieurs aliments dans chaque groupe ».

*JoëlleJuppeau, Bien manger : Les grands principes d'une bonne alimentation, Dans [www.passeportsante.net/fr](http://www.passeportsante.net/fr) publié en 2018*

N'oublie pas de : conjuguer les verbes au présent de l'indicatif, employer les connecteurs d'énumération et respecter la structure du texte explicatif.

**Les critères de correction**, que contient la grille de correction, sont à communiquer aux apprenants.

**Situation 03**

<b>Compétence évaluée</b>	
Comprendre et produire des textes explicatifs en tenant compte des contraintes de la situation de communication	
<b>Épreuve</b>	
<b>Établissement</b> : ..... <b>Type</b> : ..... <b>Matière</b> : ..... <b>Niveau</b> : ..... <b>Durée</b> : .....	
<b>Évaluation</b>	
<p>Ton frère aîné, qui joue au football, a des entraînements réguliers. Quand il rentre à la maison, il prend une douche chaude puis il se repose. Mais, le chauffe-eau a cessé de fonctionner et sa réparation prendra quelques jours. Constant que ton frère réfléchit à une solution, tu lui suggères de prendre des douches froides. Ta proposition est rejetée par tous les membres de ta famille. Surpris de leur réaction, tu décides de collecter des informations sur les avantages de la douche froide pour les publier dans la revue Santé de ton établissement. Pour rédiger tes textes, tu dois lire les ressources suivantes pour :</p> <ul style="list-style-type: none"> <li>- Définir la douche froide (un texte de 05 lignes).</li> <li>- En 08 lignes, expliquer par écrit les avantages de la douche froide.</li> <li>- Expliquer en 08 lignes pourquoi la douche froide ne peut pas remplacer la douche chaude.</li> </ul>	
<b>Document 1</b> <p>« L'eau est source de vie. Elle représente un grand pourcentage de notre corps. Il faut boire deux litres d'eau par jour pour rester hydratés, et aussi, pour bénéficier de ses bienfaits, non seulement internes, mais aussi externes.</p> <p>Nous utilisons rarement l'eau froide pour nous laver, même en été. Au contraire, nous prenons des bains chauds en hiver, et tièdes pendant les mois d'été.</p> <p>Pourtant, les spas et les centres d'esthétiques emploient une technique à base d'eau froide pour activer la circulation sanguine. Elle permet d'améliorer le transport des cellules dans notre corps, et redonne toute sa beauté à notre peau [...].</p> <p>Après le sport, il est conseillé de se doucher à l'eau froide pour que les muscles retrouvent leur état habituel ».</p> <p><i>Faut-il se doucher à l'eau froide ? Dans <a href="https://amelioresetasante.com">https://amelioresetasante.com</a>, publié le 6/11/2018</i></p>	<b>Document 2</b> <p>« En réalité, l'eau froide et l'eau chaude ont toutes les deux des bienfaits pour notre santé. L'important est de savoir les utiliser au bon moment, et de manière équilibrée.</p> <p>En règle générale, nous nous douchons quasi-exclusivement à l'eau chaude, parfois à une température excessive, et bien souvent pendant trop longtemps.</p> <p>Pour notre corps, la chaleur est plus agréable, car elle permet de dilater les vaisseaux sanguins et de favoriser la relaxation.</p> <p>Cependant, si nous en abusons, les douches à l'eau chaude peuvent nous affaiblir, nous voler notre énergie et déséquilibrer notre circulation sanguine.</p> <p>Pour toutes ces raisons, il est important de savoir combiner les deux températures pour se laver.</p> <p>De nombreuses stations balnéaires et autres centres d'hydrothérapie, proposent des bains à température alternée pour profiter de tous les bienfaits que peut nous apporter l'eau.</p> <p>Pour commencer, l'idéal est de vous appliquer un peu d'eau froide de manière graduelle, en la</p>

	combinant à intervalles réguliers avec de l'eau chaude ». <i>Les bienfaits d'une douche froide, Dans <a href="http://www.lebuteur.com">www.lebuteur.com</a> , publié le 11/3/2020</i>
<b>Document 3</b>	<b>Document 4</b>
<p>« Lorsqu'il fait froid, peu de choses sont plus agréables qu'une douche chaude. Le matin, elle est un réconfortant cocon dans lequel se réveiller ; le soir, elle est un moyen de se réchauffer après une journée parfois difficile. Pourtant, puisque rien n'est jamais parfait, les douches chaudes ont un défaut : elles peuvent assécher la peau et les cheveux [...]. Garder sa peau hydratée ne veut pas forcément dire s'infliger systématiquement des douches froides, même si celles-ci sont évidemment meilleures pour la peau. Une solution plus simple existe pour continuer à passer du temps sous l'eau chaude tout en protégeant votre peau : il s'agit de finir votre douche chaude par un trait d'eau froide, comme l'on finit certains cocktails par un trait d'absinthe. Bien entendu, essayez, pour plus d'efficacité, de passer le moins de temps possible sous l'eau chaude. De cette manière, votre peau n'aura pas le temps de s'assécher ».</p> <p><i>Matthieu Morge-Zocconi, Et s'il est temps d'abandonner votre douche chaude ? Dans <a href="https://horace.co/fr">https://horace.co/fr</a> , publié le 4/9/2017</i></p>	<p>« Le stress, tout le monde connaît ! Et des méthodes pour y remédier, on en cherche toujours et on a du mal à en trouver qui soient simples et efficaces. Avec mon truc, de l'eau froide suffit ».</p> <p><i>Céline P., Comment lutter contre le stress avec de l'eau froide ? Dans <a href="http://www.comment-economiser.fr">www.comment-economiser.fr</a> , publié le 30/9/2013</i></p>
N'oublie pas de : conjuguer les verbes au présent de l'indicatif, employer les connecteurs d'énumération et respecter la structure du texte explicatif.	
<b>Les critères de correction</b> , que contient la grille de correction, sont à communiquer aux apprenants.	

**Situation 04**

Compétence évaluée
Comprendre et produire des textes explicatifs en tenant compte des contraintes de la situation de communication
Epreuve
<b>Etablissement</b> : .....
<b>Type</b> : .....
<b>Matière</b> : .....
<b>Niveau</b> : .....
<b>Durée</b> : .....
Evaluation
Dans ton quartier, une nouvelle salle de sport vient d'être ouverte. Comme des abonnements sont ouverts à tous, tu décides de t'inscrire pour y pratiquer la boxe. Mais pour finaliser ton dossier, il te faut l'autorisation de ton père. Rentré à la maison, ton père manifeste un refus catégorique. Il t'explique en effet que les activités physiques ne se réduisent pas à celles pratiquées dans les salles de sport. Au lieu de la boxe, ton père te propose la marche à pied. Convaincu de de son idée, tu décides

de publier sur ta page Facebook trois textes -de huit (08) lignes- que tu adresses à tes amis pour les inviter à pratiquer cette activité.

Dans ces textes, que tu rédiges en t'appuyant sur les documents ci-dessous présentés, tu dois répondre aux questions qui suivent :

- Pourquoi faut-il pratiquer une activité physique ?
- Quels sont les avantages de la marche à pied ?
- Comment la marche à pied lutte-t-elle contre l'obésité ?

### Document 1



« Pour qu'une activité physique soit efficace, elle doit durer. Pour cela, mieux vaut dès le départ faire le choix d'une activité qui vous plaît. Son intégration dans votre vie au quotidien sera d'autant plus facile. Si vous n'avez pas été actif pendant un certain temps, préférez une activité peu intense. L'objectif est d'habituer votre corps à cette nouvelle activité ».

*Quels sports pour commencer ? Dans [www.dinnosante.fr](http://www.dinnosante.fr), publié le 19/05/2014*

### Document 2

«La sédentarité, comme facteur de risque collectif, est passée devant le tabagisme car elle touche davantage de monde. Le fait d'être sédentaire augmente les facteurs de risque cardiovasculaires classiques, comme le surpoids, le diabète, l'hypertension artérielle, les dyslipidémies», poursuit le Pr Galinier. Fait inquiétant, «la jeune génération est à haut risque car le temps d'activité sportive a fortement régressé ces dernières années. Il y a davantage d'enfants qui ne font rien, la facilité étant d'allumer la télévision, un écran d'ordinateur et pire, de grignoter dans le même temps».

*Nathalie Szapiro-Manoukian, La sédentarité, une menace silencieuse pour notre santé, Dans <https://sante.lefigaro.fr>, publié le 22/1/2018*

### Document 3



« Beaucoup plus qu'une activité... un sport  
Même si la marche n'est pas une activité aussi intense que la course, elle n'en demeure pas moins un sport complet. En plus de vous aider à maintenir une bonne forme physique, la marche permet de faire travailler les articulations, les muscles, les tendons ainsi que le système cardiorespiratoire. Il est évident que les membres inférieurs sont les plus sollicités, mais les autres parties du corps sont également stimulées pendant l'effort.  
Ses effets sur votre forme physique, mais aussi sur votre corps, peuvent être visibles à condition que vous la pratiquiez quotidiennement ».

*Caroline Drolet, La marche à pieds : bienfaits et bonnes pratiques, Dans <https://oppq.qc.ca>, consulté en 2020*

### Document 4



« Mais marcher est bien plus que ça, la marche est un sport à part entière et un excellent moyen de brûler des calories et de ne pas être sédentaire.  
Une simple promenade quotidienne permet de rester actif et de maintenir sa condition physique, peu importa l'âge.  
Bien entendu, pour tirer un maximum de bénéfices de la marche à pied, rien ne vaut une bonne marche soutenue, un type d'exercice que l'on considère comme étant d'intensité modérée ».

*Marche à pieds, Dans [www.sport-equipements.fr](http://www.sport-equipements.fr), consulté en 2020*

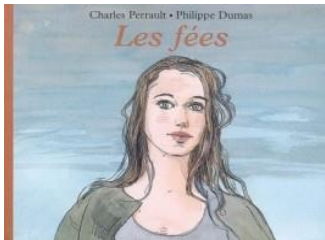


Il est important de : conjuguer les verbes au présent de l'indicatif, employer les connecteurs d'énumération et respecter la structure du texte explicatif.

Les critères de correction, que contient la grille de correction, sont à communiquer aux apprenants.

## 6.2. Deuxième année moyenne

### Situation 01

<b>Compétence évaluée</b>	
Comprendre et produire des récits de fiction en tenant compte des contraintes de la situation de communication <sup>120</sup>	
<b>Épreuve</b>	
<b>Établissement</b> : ..... <b>Type</b> : ..... <b>Matière</b> : ..... <b>Niveau</b> : ..... <b>Durée</b> : .....	
<b>Évaluation</b>	
<p>Le concours « <i>Jeunes talents</i> », qui choisit le meilleur conte merveilleux, est lancé dans les établissements de ta région. Pour obtenir ce prix, les participants doivent créer de nouveaux contes à partir des contes connus. Intéressé (e) par l'idée de ce concours, tu décides d'y participer. Les organisateurs ont sélectionné le conte étranger <i>Les fées</i> que de nombreux participants ne connaissent pas ; ce qui les a obligés à sélectionner les documents suivants, qui fournissent des précisions sur les textes à rédiger. En exploitant ces documents, tu dois écrire trois textes de 08 lignes pour :</p> <ul style="list-style-type: none"> <li>- Raconter le conte proposé en respectant son schéma narratif.</li> <li>- Faire le portrait physique et moral de chacune des deux sœurs.</li> <li>- Insérer des dialogues entre les personnages du conte.</li> </ul> <p>Les textes rédigés seront évalués puis présentés à un public constitué des élèves des établissements, des enseignants et des parents des participants.</p>	
<b>Document 1</b> 	<b>Document 2</b> <p>« <b>Schéma narratif</b>  <b>Situation initiale</b> : une jeune fille vit avec sa sœur aînée et sa mère qui sont toutes les deux méchantes avec elle. Cette situation est typique du conte car le héros connaît un problème qu'il va devoir résoudre.  <b>Élément perturbateur</b> : la fée, la rencontre de la cadette avec la pauvre femme. On sait tout de suite qu'il s'agit d'un conte merveilleux.</p>

<sup>120</sup> Les compétences de compréhension et de production indissociablement présentées dans notre étude sont séparées dans les documents d'accompagnement consultés (2016).

<p>« Un conte vraiment intéressant. On voit bien comment la malhonnêteté des gens peut amener à la déchéance la plus totale. Les gens généreux et honnêtes sont toujours récompensés pour leur vertu mais il n'en est pas de même pour les autres. Ainsi au travers de ce conte, l'auteur compare ces deux sœurs en les exposant à des situations différentes pour voir comment elles réagiraient. Elles reçurent donc ce qu'elles méritaient ! C'est une très belle leçon de vie et ceux qui le méritent peuvent enfin être heureux. Certes la réalité n'est pas totale dans ce conte puisque l'amour n'arrive pas aussi vite, mais c'est quand même une histoire rapide et agréable à lire ! »</p> <p><i>Elana, Les Fées : Commentaires, Dans <a href="https://booknode.com">https://booknode.com</a>, consulté en 2020</i></p>	<p><b>Péripiétés :</b>  1<sup>ère</sup> : puisque la cadette lui a donné de l'eau avec gentillesse, la fée lui donne un don.  2<sup>e</sup> : quand la mère s'aperçoit du don de sa fille cadette, elle décide d'envoyer l'aînée à la fontaine.  3<sup>e</sup> : l'aînée, elle, reçoit un sort à cause de sa malhonnêteté et de son arrogance.  4<sup>e</sup> : quand la mère découvre le sort de sa fille préférée, elle veut battre la cadette qui s'enfuit.  5<sup>e</sup> : la cadette rencontre un prince qui devient amoureux d'elle.  <b>Situation finale :</b> la cadette épouse le prince alors que l'aînée va mourir au coin d'un bois.  C'est une fin heureuse pour la cadette car c'est un conte merveilleux, et les rôles se sont inversés ».</p> <p><i>Agathe, Charles Perrault : Les Fées, Dans <a href="http://www.superprof.com">www.superprof.com</a>, publié en 2007</i></p>
<p style="text-align: center;"><b>Document 3</b></p> <p><b>Les personnages</b></p> <hr/> <p>« La mère : désagréable et orgueilleuse.  Sa fille aînée, Fanchon : le portrait de sa mère, jalouse de sa sœur.  Sa fille cadette : douce, honnête et belle.  Une pauvre femme : déguisement d'une fée.  Une dame bien habillée : déguisement de la même fée.  Le fils du roi : Il tombe amoureux de la cadette ».</p> <p><i>Vikidia, Les Fées, Dans <a href="https://fr.wikidia.org">https://fr.wikidia.org</a>, consulté en 2020</i></p>	<p style="text-align: center;"><b>Document 4</b></p> <p>« <b>La fée (ou « les fées »)</b> : elle apparaît à deux reprises dans le récit sous deux aspects différents : la première fois à la sœur cadette sous forme d'une pauvre femme, la seconde fois à l'aînée, sous forme d'une princesse richement vêtue. Elle demande à chacune de lui donner à boire. Elle est « omnisciente », elle connaît le bon caractère de la sœur cadette et la façon dont elle est traitée et le mauvais caractère de l'aînée ».</p> <p><i>Robin Guilloux, Charles Perrault : Les Fées, Dans <a href="https://lechatsurmonepaule.over-blog.fr">https://lechatsurmonepaule.over-blog.fr</a>, publié le 1/4/2011</i></p>
<p>Dans tes textes, il faut : employer les temps du récit, investir des connecteurs logiques et chronologiques, investir une ponctuation appropriée et respecter la structure du texte narratif.</p>	
<p><b>Les critères de correction</b>, que contient la grille de correction, sont à communiquer aux apprenants.</p>	



### Situation 02

<b>Compétence évaluée</b>
Comprendre et produire des récits de fiction en tenant compte des contraintes de la situation de communication
<b>Épreuve</b>
Établissement : .....
Type : .....
Matière : .....
Niveau : .....
Durée : .....
<b>Évaluation</b>

La revue numérique du club littéraire de ton établissement a ajouté une nouvelle rubrique *Éducation* où des textes à visée éducative seront publiés. Comme les contributions sont rares, les élèves de 2AM sont sollicités pour créer des fables et des histoires éducatives qui ne se réduisent pas à des narrations amusantes. A une semaine de la parution du nouveau numéro, ton enseignant de français, chargé de la réception et de la correction des travaux, ne reçoit aucune proposition. Pendant la séance de grammaire, il vous a soumis des documents qui incluent des informations sur la fable intitulée *le Scorpion et la Grenouille*, et vous a demandé d'écrire trois textes de 10 lignes pour :

- Présenter les personnages principaux de cette histoire.
- Rédiger des dialogues entre ces personnages.
- Réécrire cette histoire de façon à en changer la fin.

Motivé par cette activité, tu t'appuies sur les ressources présentées ci-dessous pour réaliser les tâches prescrites par ton enseignant.


<p style="text-align: center;"><b>Document 1</b></p> <p>« <i>Le Scorpion et la Grenouille</i> est une fable dans laquelle un scorpion demande à une grenouille de le transporter sur l'autre rive d'une rivière. D'abord effrayée par son aiguillon venimeux, la grenouille accepte cependant, puisque la piquer les conduirait tous deux à leur perte. Au milieu de la rivière, pourtant, le scorpion la pique mortellement. Lorsque la grenouille demande au scorpion la raison de son geste, ce dernier répond que « c'est dans sa nature. » La fable illustre le fait que certains comportements sont irrépressibles, indépendamment des conséquences ».</p> <p><i>Wikipédia, Le scorpion et la grenouille, Dans <a href="https://fr.wikipedia.org/wiki">https://fr.wikipedia.org/wiki</a>, consulté en 2020</i></p>	<p style="text-align: center;"><b>Document 2</b></p> <p>« Le scorpion a dû attendre d'être arrivé sur la terre ferme de l'autre rive et d'être descendu du dos de la grenouille pour piquer la pauvre innocente. En effet, s'il est suffisamment malin pour parler à une grenouille, il doit l'être pour dominer quelque peu sa nature et attendre pour piquer d'être en sécurité totale ».</p> <p><i>Yves Morin, Fable sur la sagesse : la grenouille et le scorpion, Dans <a href="https://groups.google.com">https://groups.google.com</a>, consulté en 2020</i></p>
<p style="text-align: center;"><b>Document 3</b></p> <p>« Poursuivis par le feu, une grenouille et un scorpion se retrouvent devant un cours d'eau, ultime rempart à franchir. Le scorpion demande à la grenouille de le prendre sur son dos pour la traversée. De crainte qu'il ne la pique, celle-ci refuse d'abord... ».</p>  <p><i>Laurent Vercueil, Et le scorpion ne piqua pas la grenouille, Dans <a href="http://www.echosciences-grenoble.fr">www.echosciences-grenoble.fr</a>, publié le 7/7/2016</i></p>	<p style="text-align: center;"><b>Document 4</b></p> <p>« <i>Moralité</i> : quand vous êtes auprès d'un scorpion. Quel que soit son discours ou ses bonnes intentions, soyez sur vos gardes, ne le croyez pas. A la première occasion il ne pourra pas s'empêcher de suivre sa nature profonde et vous en paierez les conséquences ».</p>  <p><i>Laurent Combaud, La fable de la grenouille et du scorpion, Blog de l'auteur, publié le 8/8/2008</i></p>

Dans tes textes, tu dois : employer les temps du récit, investir des connecteurs logiques et chronologiques, investir une ponctuation appropriée et respecter la structure du texte narratif.

**Les critères de correction**, que contient la grille de correction, sont à communiquer aux apprenants.

### Situation 03

<b>Compétence évaluée</b>	
Comprendre et produire des récits de fiction en tenant compte des contraintes de la situation de communication	
<b>Épreuve</b>	
<b>Établissement</b> : ..... <b>Type</b> : ..... <b>Matière</b> : ..... <b>Niveau</b> : ..... <b>Durée</b> : .....	
<b>Évaluation</b>	
<p>A cause de la pandémie et du confinement, les enfants de ta ville -habitué(e)s à une vie libre- se retrouvent emprisonnés dans les maisons. Leurs parents organisent une réunion -à laquelle tu assistes- et proposent des sorties vers des sites touristiques. Cependant, cette décision n'est pas approuvée car la distanciation sociale ne pourrait pas être assurée. Tu leur proposes alors de créer des représentations qui seront données dans le théâtre où la distanciation serait garantie. Ta proposition est applaudie par les parents et la mission de créer la première pièce théâtrale t'est attribuée. Comme le spectacle est destiné aux enfants, tu optes pour la dramatisation de la fable le corbeau et le renard. Tu dois alors écrire trois textes pour :</p> <ul style="list-style-type: none"> <li>- Présenter les personnages et les décrire physiquement et moralement.</li> <li>- Rédiger les répliques du renard qui tente de convaincre le corbeau de parler.</li> <li>- Propose une fin différente à cette histoire.</li> </ul> <p>Pour la réalisation de ces tâches qui aboutissent à des textes écrits, d'une longueur de (10) dix lignes, destinés aux membres-acteurs de ton groupe, tu dois t'appuyer sur les documents suivants.</p>	
<b>Document 1</b>	<b>Document 2</b>
<p>« <b>Le Corbeau et le Renard</b>  <i>Jean de La Fontaine</i>  <b>Résumé de l'histoire</b> : Un corbeau fier et orgueilleux perché sur un arbre et un renard rusé et flatteur qui veut lui prendre son fromage...  <b>Explication de la morale</b> : Il faut savoir garder la raison même quand quelqu'un nous flatte et nous dit ce que l'on veut entendre. La sagesse est toujours la plus forte ».</p> <p><i>LAROUSSE, Le corbeau et le renard, Informations sur le livre, consulté en 2020</i></p>	<p>Un corbeau, ayant volé un morceau de viande, s'était perché sur un arbre. Un renard l'aperçut, et, voulant se rendre maître de la viande, se posta devant lui et loua ses proportions élégantes et sa beauté, ajoutant que nul n'était mieux fait que lui pour être le roi des oiseaux, et qu'il le serait devenu sûrement, s'il avait de la voix. Le corbeau, voulant lui montrer que la voix non plus ne lui manquait pas, lâcha la viande et poussa de grands cris. Le renard se précipita et, saisissant le morceau, dit : « Ô corbeau, si tu avais aussi du jugement, il ne te manquerait rien pour devenir le roi des oiseaux. »</p> <p>Cette fable est une leçon pour les sots.</p> <p><i>Esope, Le corbeau et le renard, Dans <a href="https://fr.wikisource.org">https://fr.wikisource.org</a>, 1927</i></p>
<b>Document 3</b>	<b>Document 4</b>
<p>« Maître Corbeau, sur un arbre perché,          Tenait en son bec un fromage.</p>	<p>« Le renard est un personnage qui revient à plusieurs reprises dans les fables de La Fontaine. Il est</p>

<p>Maître Renard, par l'odeur <b>alléché</b>, Lui tint à peu près ce langage : Et bonjour, Monsieur du Corbeau. Que vous êtes joli ! que vous me semblez beau ! Sans mentir, si votre <b>ramage</b> <b>Se rapporte</b> à votre plumage, Vous êtes le <b>Phénix</b> des <b>hôtes</b> de ces bois. À ces mots, le Corbeau <b>ne se sent pas de joie</b> ; Et pour montrer sa belle voix, Il ouvre un large bec, laisse tomber sa <b>proie</b>. Le Renard s'en saisit, et dit : Mon bon Monsieur, Apprenez que tout flatteur Vit aux <b>dépens</b> de celui qui l'écoute. Cette leçon vaut bien un fromage, sans doute. Le Corbeau honteux et <b>confus</b> Jura, mais un peu tard, qu'on ne l'y prendrait plus ».</p> <p style="text-align: right;"><i>Jean de La Fontaine, Fables.</i></p>	<p>traditionnellement représenté comme un personnage rusé...</p>  <p>Dans la fable « <i>Le Corbeau et le Renard</i> » de La Fontaine, le renard élabore une stratégie afin d'obtenir le fromage que tient le corbeau dans son bec. Il flatte alors le corbeau pour que celui-ci laisse tomber son fromage ».</p> <p><i>La ruse et le renard chez La Fontaine, Dans <a href="http://www.schoolmouv.fr">www.schoolmouv.fr</a>, consulté en 2020</i></p>
<p>Dans tes textes, tu dois : employer les temps du récit, investir des connecteurs logiques et chronologiques, investir une ponctuation appropriée et respecter la structure du texte narratif.</p>	
<p><b>Les critères de correction</b>, que contient la grille de correction, sont à communiquer aux apprenants.</p>	

### Situation 04

Compétence évaluée
Comprendre et produire des récits de fiction en tenant compte des contraintes de la situation de communication
Épreuve
<p><b>Établissement</b> : .....</p> <p><b>Type</b> : .....</p> <p><b>Matière</b> : .....</p> <p><b>Niveau</b> : .....</p> <p><b>Durée</b> : .....</p>
Évaluation
<p>Dans le cadre de la numérisation des écoles, ton établissement a lancé le projet de la bibliothèque numérique où les productions des apprenants dans toutes les disciplines seront publiées pour être consultées par un public plus large. Pour la langue française, les enseignants optent pour la publication des narrations qui relèvent des divers genres. La mission de rédiger des fables est confiée aux élèves de la 2AM dont tu fais partie. Ton enseignant te propose la fable le Lièvre et la Tortue que tu ne connais pas. Tu as alors posté une question dans le forum <i>Histoires</i> pour obtenir des informations sur cette fable. Les réponses formulées par des internautes, représentées dans les documents suivants, doivent être lues et comprises pour :</p> <ul style="list-style-type: none"> <li>- Rédiger des dialogues entre les personnages principaux.</li> <li>- Insérer des descriptions de ces personnages.</li> <li>- Proposer la suite de cette fable, où le lièvre défie la tortue dans une nouvelle course.</li> </ul>

<p style="text-align: center;"><b>Document 1</b></p> <p><b>« La Tortue et le Lièvre</b> Le Lièvre considérant la Tortue qui marchait d'un pas tardif, et qui ne se traînait qu'avec peine, se mit à se moquer d'elle et de sa lenteur. La Tortue n'entendit point raillerie, et lui dit d'un ton aigre, qu'elle le défiait, et qu'elle le vaincrait à la course, quoiqu'il se vantât fièrement de sa légèreté. Le Lièvre accepta le défi. Ils convinrent ensemble du lieu où ils devaient courir, et du terme de leur course. Le Renard fut choisi par les deux parties pour juger ce différend. La Tortue se mit en chemin, et le Lièvre à dormir, croyant avoir toujours du temps de reste pour atteindre la Tortue, et pour arriver au but avant elle. Mais enfin elle se rendit au but avant que le Lièvre fut éveillé. Sa nonchalance l'exposa aux railleries des autres Animaux. Le Renard, en Juge équitable, donna le prix de la course à la Tortue.</p> <p><i>Esopé (VIIe-VIe siècle av. J.-C)</i></p>	<p style="text-align: center;"><b>Document 2</b></p> <p style="text-align: center;"><b>« LA TORTUE ET LE LIÈVRE</b></p> <p>La tortue et le lièvre disputaient qui était le plus vite. En conséquence ils fixèrent un jour et un endroit et se séparèrent. Or le lièvre, confiant dans sa vitesse naturelle, ne se pressa pas de partir ; il se coucha au bord de la route et s'endormit ; mais la tortue, qui avait conscience de sa lenteur, ne cessa de courir, et, prenant ainsi l'avance sur le lièvre endormi, elle arriva au but et gagna le prix ».</p> <p><i>Esopé, La tortue et le lièvre, Dans <a href="https://fr.wikisource.org">https://fr.wikisource.org</a>, 1927</i></p>
<p style="text-align: center;"><b>Document 3</b></p> <p><b>« Dès le début de la fable, la vanité du lièvre apparaît</b> car il demande à la tortue si elle n'a pas perdu la raison. Finalement c'est elle qui gagne la course, montrant ainsi que ses propos étaient censés, contrairement à ceux du lièvre. Le lièvre fait preuve de vanité en vaquant à toute sorte d'occupations inutiles plutôt que de se soucier de l'avance de la tortue. <b>On retrouve ici la critique de la vanité chère à la Fontaine</b> ».</p> <p><i>Le lièvre et la tortue, Dans <a href="https://bacdefrancais.net">https://bacdefrancais.net</a>, consulté en 2020</i></p>	<p style="text-align: center;"><b>Document 4</b></p> <p><b>« Et si on continuait l'histoire ?</b> Le lièvre, déçu après avoir perdu, a fait un examen de conscience et a reconnu ses erreurs. Il a découvert qu'il avait perdu la course à cause de son orgueil, pour être aussi prétentieux et négligeant. S'il n'avait pas pensé que c'était gagné d'avance, La tortue ne l'aurait jamais vaincu. – Alors, il défia la tortue à une nouvelle course ». <i>Le lièvre et la tortue : morale et suite, Dans <a href="http://www.femmes-tendances.com">www.femmes-tendances.com</a>, publié en 2019</i></p>
<p>Dans tes trois productions dont la longueur est d'environ dix (10) lignes, tu dois : employer les temps du récit, investir des connecteurs logiques et chronologiques, investir une ponctuation appropriée et respecter la structure du texte narratif.</p>	
<p><b>Les critères de correction</b>, que contient la grille de correction, sont à communiquer aux apprenants.</p>	

### 6.3. Troisième année moyenne

#### Situation 01

<b>Compétence évaluée</b>
Comprendre et produire <b>des récits de faits réels</b> en tenant compte des contraintes de la situation de communication <sup>121</sup>
<b>Épreuve</b>
<b>Établissement</b> : .....
<b>Type</b> : .....

<sup>121</sup> Les compétences de compréhension et de production indissociablement présentées dans notre étude sont séparées dans les documents d'accompagnement consultés (2016).

<b>Matière</b> : .....	
<b>Niveau</b> : .....	
<b>Durée</b> : .....	
<b>Évaluation</b>	
<p>A l'occasion de la journée du Savoir, qui correspond au 16 avril, ton établissement organise une manifestation culturelle : des reportages réalisés par les apprenants seront diffusés, des affiches représentant des événements historiques seront exposées... Ces projets réalisés par les apprenants seront par la suite publiés dans le site de la Direction de l'Éducation. Ne pouvant pas réaliser un reportage vu les moyens que ce projet-ci requiert, tu optes pour le second projet : réaliser une affiche sur les événements du <i>20 août 1955</i>. Quand les tâches sont réparties, les tiennes consistent à :</p> <ul style="list-style-type: none"> <li>- Raconter ce qui s'est passé ce jour-là.</li> <li>- Expliquer l'importance de ces événements dans la guerre de libération.</li> <li>- Préciser les conséquences de ces événements sur les Algériens.</li> </ul> <p>Pour rédiger ces textes, dont la longueur est d'environ dix (10) lignes, tu dois lire les ressources suivantes collectées et repérer les informations utiles.</p>	
<b>Document 1</b>	<b>Document 2</b>
<p>« Intervenant dans un contexte marqué par une répression aveugle du colonisateur français envers la population et les moudjahidine retranchés dans la région des Aurès, dans l'Est du pays, les attaques du 20 août 1955 du Nord-constantinois, dont l'un des objectifs était de desserrer l'étau sur cette région, a insufflé un second souffle à la Révolution naissante de Novembre 1954.</p> <p>Orchestrées sous l'impulsion de Zighoud Youcef, commandant de la Wilaya II historique après la mort de Didouche Mourad le 18 janvier 1955, les opérations menées dans la région de Constantine-Skikda-Guelma, représentent "une date phare dans l'Histoire de la Révolution pour la libération du pays," se souvient encore Mohamed-Salah Lakher, 88 ans, membre des premiers groupes de choc de fidaï et militant actif de l'organisation du FLN à Constantine ».</p> <p><i>20 août 55 à Constantine : les opérations des fidaï propulsent la Révolution, Dans <a href="http://www.aps.dz">www.aps.dz</a>, publié le 18/8/2020</i></p>	<p>« L'offensive du 20 août 1955 dans le Nord-Constantinois est «le véritable déclenchement de la Révolution», estime Rabah Lounici, chercheur en histoire. «Zighout Youcef a démontré ses capacités de meneur et de leader de ses troupes pour briser l'étau qui s'est resserré sur la région de l'Aurès totalement encerclée par les forces coloniales», souligne le chercheur, estimant que parmi les autres facteurs ayant poussé ce leader à organiser cette offensive, «la volonté de remonter le moral de la population qui hésitait encore à embrasser la Révolution».</p> <p><i>Fatima Ait Khaldoun, Offensive du 20 août 1955 : un tournant décisif pour la Révolution, Dans <a href="http://www.elwatan.com">www.elwatan.com</a>, publié le 20/8/2018</i></p>

Document 3	Document 4
<p>« La spectaculaire offensive menée le 20 août 1955 par Zighoud Youcef dans le Nord-Constantinois et sauvagement réprimée dans le sang aura été un tournant décisif dans le parcours de la Révolution algérienne, en même temps qu'elle a illustré la nature criminelle de la puissance coloniale française, selon des témoins et autres historiens et chercheurs.</p> <p>Vaillamment orchestrée en plein jour par le Commandant de la Zone II (Nord-Constantinois) et son adjoint Lakhdar Bentobbal, cette attaque, d'une grande ampleur, avait ciblé plusieurs objectifs de l'occupant, principalement dans les villes de Constantine, Skikda, Guelma et Collo ».</p> <p><i>Révolution algérienne : les offensives du 20 août 55, un acte sauvagement réprimé dans le sang, Dans <a href="http://www.aps.dz">www.aps.dz</a>, publié le 31/5/2020</i></p>	<p>« Deux massacres d'Européens ont bien été perpétrés par les insurgés dans deux villages du Constantinois : à El Alia, 35 villageois ont été tués, et 7 villageois à Ain Abid. Les 29 autres victimes comptabilisées ont péri en divers endroits, dans des circonstances rarement précisées, lors d'une succession de crimes mais "non d'un crime de masse ", précise M<sup>me</sup> Mauss-Copeaux. Le tableau faisant état d'un massacre généralisé des Européens dans le Constantinois, tel qu'il a souvent été rapporté par les médias français, est donc sans rapport avec la réalité ».</p> <p><i>Florence Beaugé, « Algérie : 20 août 1955. Insurrection, répression, massacre » de Claire Mauss-Copeaux : répression démesurée, Dans <a href="http://www.lemonde.fr">www.lemonde.fr</a>, publié le 28 février 2011</i></p>
<p>Dans tes productions, il est important de : respecter les temps du récit, employer les connecteurs logiques et chronologiques, investir une ponctuation appropriée et respecter la structure des textes narratifs.</p>	
<p><b>Les critères de correction</b>, que contient la grille de correction, sont à communiquer aux apprenants.</p>	

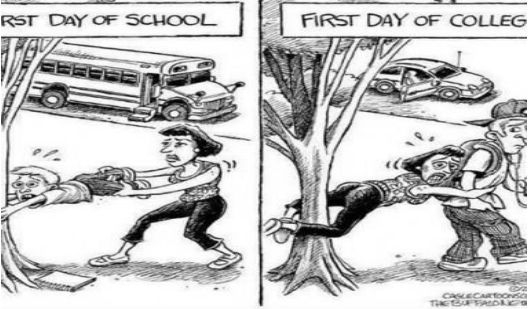
### Situation 02

Compétence évaluée
<p>Comprendre et produire <b>des récits de faits réels</b> en tenant compte des contraintes de la situation de communication.</p>
Épreuve
<p><b>Établissement</b> : .....</p>
<p><b>Type</b> : .....</p>
<p><b>Matière</b> : .....</p>
<p><b>Niveau</b> : .....</p>
<p><b>Durée</b> : .....</p>
Évaluation
<p>Tu as créé un blog personnel <i>Mémoires</i> où tu publieras tes souvenirs. Tes textes seront lus et commentés par les internautes qui accèdent à ta page. Comme première étape, tu décides de publier les souvenirs qui ont marqué ton enfance. Cependant, te rappeler certains détails paraît difficile. Tu décides de te servir des souvenirs publiés par les internautes pour rédiger des textes sur ton premier jour à l'école. En t'aidant des extraits retrouvés dans des sites internet, écris trois textes de 10 lignes pour :</p> <ul style="list-style-type: none"> <li>- Raconter les différents moments de ce souvenir (les préparations précédant la rentrée, le premier jour, le premier contact avec ton enseignant...).</li> </ul>



- Décrire tes sentiments à tous les moments de ce jour (à la maison, à l'école, dans la classe...).
- Décrire les sentiments des membres de ta famille (ta mère, ton père...).

Les ressources suivantes sont à lire pour que les tâches soient réalisées.

<p style="text-align: center;"><b>Document 1</b></p> <p>« Depuis que je suis toute petite, je rêve d'aller à l'école. Mon frère et ma sœur s'y rendent tous les jours. Je les regarde partir avec leur sac sur le dos. J'aimerais tant les suivre. Ma mère dit que je suis encore trop jeune pour aller à l'école. Je pleure, alors maman sèche mes larmes et déclare : Promis, un jour tu iras à l'école ! ».</p> <p>Mon frère et ma sœur apprennent les lettres et les chiffres. Depuis qu'ils savent lire, ils me racontent de belles histoires. Des histoires dans des livres avec des images de toutes les couleurs.</p> <p>En attendant, je joue avec mes amis à la garderie. Je sais compter jusqu'à dix et je connais toutes les lettres par cœur. Mais moi, j'ai hâte de lire des histoires dans les beaux livres colorés. Et j'aimerais connaître tous les secrets des professeurs. Papa me rassure alors : « Promis, un jour tu iras à l'école ! ». L'été est fini et les enfants se préparent pour la rentrée. Maman m'annonce que demain sera mon premier jour d'école ! Mon cœur bat très fort. Je vais enfin apprendre à lire ! »</p> <p><i>Louise Tendreau-Levert, Premier jour d'école, Dans <a href="https://naitreetgrandir.com">https://naitreetgrandir.com</a>, consulté en 2020</i></p>	<p style="text-align: center;"><b>Document 3</b></p> <p>« <b>La rentrée pour les parents</b> Le premier jour de classe ne concerne pas seulement les enfants : c'est aussi <b>la rentrée des parents</b>. La rentrée pour eux signifie fini les habitudes quotidiennes : planning adapté aux programmes des classes des enfants. Il faut les préparer le matin, puis les conduire à l'école, les chercher à la sortie des classes en fin de matinée. Quel stress ! »</p>  <p><i>Marcel, Comment était votre toute première journée d'école, Dans <a href="https://decryptages.mondoblog.org">https://decryptages.mondoblog.org</a>, publié 9/9/2014</i></p>
<p style="text-align: center;"><b>Document 3</b></p> <p>« <b>Mon premier jour d'école</b> Mon premier jour, je ne l'ai pas oublié en tout cas. J'avais de belles chaussures neuves, un beau cartable, un magnifique tablier (vu que j'allais à la première section de la maternelle de l'école publique). Eh, j'oubliais un détail, la <b>séparation fut difficile</b>, très difficile même. J'avais écrasé une larme en voyant ma mère s'éloigner progressivement de l'école ».</p> <p><i>Marcel, Comment était votre toute première journée d'école, Dans <a href="https://decryptages.mondoblog.org">https://decryptages.mondoblog.org</a>, publié 9/9/2014</i></p>	<p style="text-align: center;"><b>Document 4</b></p> <p>« La rentrée signifie vêtements neufs, fournitures scolaires de qualité, l'occasion de se faire des amis et d'acquérir des savoirs. L'expérience <b>du premier jour d'école est unique pour chacun</b>. Quitter le cocon familial pour cet environnement stimulant et parfois frustrant peut se transformer en un réel cauchemar qui marquera toute la vie, mais surtout conditionnera tout le parcours scolaire de l'enfant ».</p> <p><i>Marcel, Comment était votre toute première journée d'école, Dans <a href="https://decryptages.mondoblog.org">https://decryptages.mondoblog.org</a>, publié 9/9/2014</i></p>

Tu dois employer les temps du récit, investir des connecteurs logiques et chronologiques, investir une ponctuation appropriée et respecter la structure des textes narratifs.

**Les critères de correction**, que contient la grille de correction, sont à communiquer aux apprenants.

### Situation 03

<b>Compétence évaluée</b>
Comprendre et produire <b>des récits de faits réels</b> en tenant compte des contraintes de la situation de communication.
<b>Épreuve</b>


<b>Établissement</b> : ..... <b>Type</b> : ..... <b>Matière</b> : ..... <b>Niveau</b> : ..... <b>Durée</b> : .....	
<b>Évaluation</b>	
<p>Un stade pour les jeunes de ta région vient d'être construit. A une semaine de l'inauguration, des bénévoles sont sollicités par la mairie pour tout préparer avant le match d'ouverture. Comme tu t'es cassé la jambe lors de ton dernier match, tu décides de concevoir une affiche, qui sera exposée à l'entrée du stade, pour fournir des précisions sur la date du <i>08 mai 1945</i> choisie pour nommer ce stade. Pour que les précisions rédigées soient satisfaisantes, rédige trois textes pour :</p> <ul style="list-style-type: none"> <li>- Raconter les évènements qui se sont déroulés ce jour-là.</li> <li>- Décrire les sentiments et les émotions des Algériens le jour de ces manifestations.</li> <li>- Préciser leurs conséquences sur les Algériens et sur la guerre de libération nationale.</li> </ul> <p>Pour la rédaction de ces textes dont la longueur est d'environ (10) dix lignes, tu dois lire les documents ci-dessous pour en extraire des informations utiles.</p>	
<p style="text-align: center;"><b>Document 1</b></p> <p>« Le 8 mai 1945, à Sétif, les manifestations de la victoire sont marquées par la constitution d'un cortège se dirigeant vers le monument aux morts de la Grande Guerre. Des Algériens s'insèrent dans le cortège en portant un drapeau algérien ce qui provoque une intervention des forces de police (le jeune homme portant le drapeau est tué par un policier), un mouvement de panique dans la manifestation, des morts européens dans le cadre d'une émeute (peut-être une centaine et autant de blessés) engendre une intense répression contre les Algériens ».</p> <p><i>Le 8 mai 1945 : « une mémoire occultée », Dans <a href="http://www.reseau-canope.fr">www.reseau-canope.fr</a>, consulté en 2020</i></p>	<p style="text-align: center;"><b>Document 2</b></p> <p><b>« SETIF- L'emblème national arboré par le chahid Saâl Bouzid un certain mardi 8 mai 1945 à Sétif pour revendiquer l'indépendance de l'Algérie flotte à ce jour haut au musée du moudjahid de la même ville.</b></p> <p>Cet emblème occupe une place toute spéciale dans les cœurs des habitants de la région car il demeure attaché à une horrible tragédie humaine perpétrée par la machine de mort française qui avait fauché la vie à 45.000 victimes dont le seul tort a été de descendre à la rue pour revendiquer pacifiquement leur liberté et émancipation ».</p> <p><i>8 mai 1945 : histoire d'un drapeau témoin de l'un des plus grands massacres des temps modernes, Dans <a href="http://www.aps.dz">www.aps.dz</a>, publié le 6 mai 2020</i></p>

Document 3	Document 4
<p>« Le jour même où la France est libérée, elle réaffirme dans le sang sa domination coloniale en Algérie : 45.000 morts à Sétif, Guelma, Kherrata et dans tout le Constantinois. Cet évènement tragique nous touche particulièrement à Lyon, car il y a un grand nombre de personnes originaires de Sétif parmi les habitants de Lyon.</p> <p>Le 8 mai 1945 signifie la fin du nazisme. Il correspond aussi à l'un des moments les plus sanglants de l'histoire nationale. La répression colonialiste venait d'y faire ses premiers accroc face à une population farouchement déterminée à se promouvoir aux nobles idéaux de paix et d'indépendance ».</p> <p>8 mai 1945 : massacre de Sétif, Dans <a href="https://rebellyon.info">https://rebellyon.info</a>, publié le 8/5/2020</p>	<p>« Le 8 mai 1945, tandis que la France fêtait la victoire, son armée massacrait des milliers d'Algériens à Sétif et à Guelma. Ce traumatisme radicalisera irréversiblement le mouvement national. (...) les massacres du 8 mai 1945 dans les régions de Sétif et de Guelma sont considérés rétrospectivement comme le début de la guerre algérienne d'indépendance. Cet épisode appartient aux lignes de clivage liées à la conquête coloniale. La vie politique de l'Algérie, plus distincte de celle de la France au fur et à mesure que s'affirme un mouvement national, a été dominée par les déchirements résultant de cette situation ».</p> <p>Mohamed Harbi, <i>La guerre d'Algérie a commencé à Sétif</i>, Dans <a href="http://www.monde-diplomatique.fr">www.monde-diplomatique.fr</a>, publié en 2005</p>
<p>Tu dois employer les temps du récit, investir des connecteurs logiques et chronologiques, investir une ponctuation appropriée et respecter la structure des textes narratifs.</p> <p>La longueur des textes à produire : environ dix (10) lignes.</p>	
<p><b>Les critères de correction</b>, que contient la grille de correction, sont à communiquer aux apprenants.</p>	

### Situation 04

Compétence évaluée
<p>Comprendre et produire <b>des récits de faits réels</b> en tenant compte des contraintes de la situation de communication.</p>
Épreuve
<p><b>Établissement</b> : .....</p>
<p><b>Type</b> : .....</p>
<p><b>Matière</b> : .....</p>
<p><b>Niveau</b> : .....</p>
<p><b>Durée</b> : .....</p>
Évaluation
<p>Ton établissement a créé un journal électronique où des événements variés sont rapportés aux élèves. Une publication hebdomadaire des articles rédigés par les élèves de 3AM est prévue pour une meilleure sensibilisation aux dangers de la route. Cependant, seuls les articles conformes aux exigences sont sélectionnés par le responsable du projet. Motivé par cette initiative, tu décides de contribuer. Pour le premier numéro, tu dois :</p> <ul style="list-style-type: none"> <li>- Rédiger un article qui rapporte l'accident évoqué dans le témoignage.</li> <li>- Insérer un chapeau qui ne rapporte que les informations importantes.</li> <li>- Insérer deux témoignages des habitants qui ont aidé la dame.</li> </ul>

Pour ce faire, tu dois lire minutieusement les documents qui vous sont soumis par les élèves chargés de la collecte des informations.

<b>Document 1</b>	<b>Document 2</b>
 <p>« Il y a trois ans, Cynthia, 34 ans, a fauché mortellement un motard alors qu'elle était au volant de sa voiture. Même si elle a été innocentée, sa vie a changé à tout jamais. Chaque 1er février, j'aimerais ne pas me réveiller, afin de ne pas revivre mentalement cette terrible journée (silence). Il y a trois ans, mon pire cauchemar est devenu réalité... ».</p> <p><i>Barbara Wesoly, Cynthia, 34 ans a provoqué un accident mortel, Dans <a href="http://www.flair.be">www.flair.be</a>, publié le 6/12/2019</i></p>	<p>« (...) C'était un lundi comme les autres. Je travaillais pour une compagnie d'assurances et ma journée s'était terminée par un entretien d'évaluation positif. Je m'apprêtais à rentrer à la maison, quand je me suis rappelé que je devais faire le plein d'essence et je ne voulais pas prendre le risque de tomber en panne. Je me suis arrêtée à la station-service en face de mon bureau. Elle se trouve sur une route très fréquentée et nous étions en pleine heure de pointe. En quittant les lieux, j'ai regardé à gauche, à droite, puis une seconde fois à gauche, et j'ai finalement attendu qu'une conductrice me laisse passer avant de me lancer. Je me suis donc engagée sur la voie à une vitesse très limitée. Quand soudain j'ai entendu un bruit sourd.</p> <p>(...) Il fallait que je sorte, vite ! Paniquée, je suis tombée au sol. Quelques secondes plus tard, on me relevait et on m'emmenait de l'autre côté de la rue, où des habitants du quartier m'ont proposé de m'asseoir. Quand j'ai repris mes esprits, j'ai regardé en direction de ma voiture. Quelques mètres plus loin, une personne gisait sur le sol. Elle ne bougeait plus. J'étais sous le choc ».</p> <p><i>Barbara Wesoly, Cynthia, 34 ans a provoqué un accident mortel, Dans <a href="http://www.flair.be">www.flair.be</a>, publié le 6/12/2019</i></p>
<p><b>Document 3</b></p> <p>« (...) Au mois de novembre 2016, un huissier de justice a frappé à ma porte avec une mauvaise nouvelle. J'étais poursuivie pour coups et blessures ayant entraîné la mort. Je suis restée sans voix, brisée. Christophe a, une nouvelle fois, été d'un grand réconfort. Mes souvenirs sont flous, mais je pense que c'est précisément ce jour-là que j'avais décidé de partager publiquement mon histoire sur Facebook afin d'expliquer à mes amis la raison de mon absence ces derniers mois ».</p> <p><i>Barbara Wesoly, Cynthia, 34 ans a provoqué un accident mortel, Dans <a href="http://www.flair.be">www.flair.be</a>, publié le 6/12/2019</i></p>	<p><b>Document 4</b></p> <p>« (...) On m'a présenté mon dossier et, pour la première fois, j'ai obtenu des éléments de réponse sur ce qu'il s'était réellement produit le jour de l'accident. J'ai appris, par exemple, que le motard conduisait sans être assuré et qu'on avait détecté la présence d'alcool et de drogue dans son sang. Cet argument plaidait en ma faveur, même si le tribunal allait surtout vérifier si le code de la route avait été respecté ».</p> <p><i>Barbara Wesoly, Cynthia, 34 ans a provoqué un accident mortel, Dans <a href="http://www.flair.be">www.flair.be</a>, publié le 6/12/2019</i></p>

Les textes produits doivent employer les temps du récit, investir des connecteurs logiques et chronologiques, investir une ponctuation appropriée et respecter la structure des textes journalistiques.

**Les critères de correction**, que contient la grille de correction, sont à communiquer aux apprenants.

## 6.4. Quatrième année moyenne

### Situation 01

Compétence évaluée

Comprendre et produire <b>des textes argumentatifs</b> en tenant compte des contraintes de la situation de communication <sup>122</sup>	
<b>Épreuve</b>	
<b>Établissement</b> : .....	
<b>Type</b> : .....	
<b>Matière</b> : .....	
<b>Niveau</b> : .....	
<b>Durée</b> : .....	
<b>Évaluation</b>	
<p>En dépit de sa grande beauté, l'Algérie est un pays où le nombre des touristes qui débarquent annuellement demeure restreint. D'ailleurs, même les touristes algériens préfèrent fuir le pays pour passer leurs vacances dans des pays qui ne sont pas toujours aussi charmants. Pour les inciter à découvrir sa beauté et encourager par-là le tourisme en Algérie, tu proposes à tes camarades de concevoir des dépliants pour présenter certains lieux touristiques du territoire national. Après la répartition des tâches sur les membres du groupe, tu dois rédiger trois textes pour :</p> <ul style="list-style-type: none"> <li>- Présenter le pays et citer ses endroits les plus connus.</li> <li>- Expliquer pourquoi ce pays doit être visité.</li> <li>- Expliquer pourquoi il est nécessaire d'encourager le tourisme intérieur.</li> </ul> <p>Pour réaliser ces tâches, la lecture des documents suivants est nécessaire :</p>	
<b>Document 1</b>	<b>Document 2</b>
<p>« Entre les visites de sites préhistoriques et leurs peintures rupestres, les vestiges romains, les cités médiévales, l'architecture d'influence orientale ou coloniale, vous ne risquez pas de vous ennuyer pendant vos vacances en Algérie ! La lumière a influencé les peintres occidentaux au 19<sup>ème</sup> siècle (les orientalistes) et jusqu'à Paul Klee. Alger dispose de plusieurs musées, notamment du Bardo ou des Beaux-Arts. Laissez-vous envoûter par les musiques traditionnelles, notamment le Chaâbi qui allie le style classique arabo-andalou aux sonorités occidentales ou encore le Raï. Plusieurs fêtes animent le mois de juillet : festival Panafricain d'Alger ou d'Oran. La réouverture du festival de musique de Timgad est également prévue, grâce à l'édification d'une nouvelle salle pour ne plus risquer d'abîmer les arènes romaines ».</p> <p><i>Caroline, Guide touristique pour visiter l'Algérie et bien préparer son voyage, Dans <a href="http://www.alibabuy.com">www.alibabuy.com</a>, consulté en 2020</i></p>	<p>« En tout premier lieu, il faut visiter la cité romaine de Timgad qui se trouve au nord-est de l'Algérie. Longtemps ensevelie sous le sable du désert elle a été préservée de l'usure du temps, ce qui permet de visiter un site avec des maisons et des murs à hauteur d'homme ainsi qu'un ensemble composé, entre autres, d'un théâtre, d'une bibliothèque, d'un forum et de thermes. Près de Constantine se trouve un autre site romain : Djemila, tandis que Tipasa se situe sur la côte, à 70 kilomètre d'Alger. La capitale abrite un autre site remarquable ; la Casbah et ses ruelles en pente, immortalisée par le film Pépé le Moko. Profitez-en pour admirer la grande mosquée d'Alger et arpenter le quartier populaire de Bab El-Oued ».</p> <p><i>Caroline, Guide touristique pour visiter l'Algérie et bien préparer son voyage, Dans <a href="http://www.alibabuy.com">www.alibabuy.com</a>, consulté en 2020</i></p>
<b>Document 3</b>	<b>Document 4</b>

<sup>122</sup> Les compétences de compréhension et de production indissociablement présentées dans notre étude sont séparées dans les documents d'accompagnement consultés (2016).

<p>« Salut, Comme les précédents posts, je conseille à tous de visiter l'Algérie. Je me trouve d'ailleurs en ce moment à Alger, et ce pendant encore une dizaine de jours dans le cadre de la mission Djeet organisée par mon école. Je conseille notamment :</p> <ul style="list-style-type: none"> <li>- d'aller dans le sud (à Taghit) pour y voir les ergs sahariens,</li> <li>- d'aller sur la côte qui est magnifique du côté de Tipaza,</li> <li>- et de rencontrer et d'échanger avec le maximum d'algériens, des gens super cools.</li> </ul> <p>Alors bon voyage aux futurs visiteurs. »</p> <p><i>Rikko, Les étrangers ont eu une mauvaise image de l'Algérie, Dans <a href="http://www.tripadvisor.fr">www.tripadvisor.fr</a>, publié en 2008</i></p>	<p>« Mis à part les risques terroristes, les infrastructures hormis quelques-unes ne se prêtent pas au tourisme. Lorsqu'un hôtel affiche 5 étoiles quand il n'en vaut pas plus de 2, voire 1 rien ne donne envie d'y aller. Le touriste s'y sent floué. Je ne parle pas des grands hôtels d'Alger. Sinon il n'y a aucune fausse idée ni préjugé d'aucune sorte. Certains préférant le sud s'y rendent régulièrement bien que le ministère des affaires étrangères le déconseille formellement. L'accueil y est plutôt amical, les paysages magnifiques.</p> <p><i>Andrea S, Les étrangers ont eu une mauvaise image de l'Algérie, Dans <a href="http://www.tripadvisor.fr">www.tripadvisor.fr</a>, publié en 2008</i></p>
<p>Dans tes productions, il est important de mettre les verbes au présent de l'indicatif, employer les connecteurs logiques et chronologiques, investir une ponctuation appropriée et respecter la structure des textes à visée argumentative.</p> <p>Longueur des textes à produire : environ douze (12) lignes.</p>	
<p><b>Les critères de correction</b>, que contient la grille de correction, sont à communiquer aux apprenants.</p>	

### Situation 02

<b>Compétence évaluée</b>
Comprendre et produire <b>des textes argumentatifs</b> en tenant compte des contraintes de la situation de communication <sup>123</sup>
<b>Épreuve</b>
<p><b>Établissement</b> : .....</p> <p><b>Type</b> : .....</p> <p><b>Matière</b> : .....</p> <p><b>Niveau</b> : .....</p> <p><b>Durée</b> : .....</p>
<b>Évaluation</b>
<p>Un documentaire impressionnant était diffusé sur la chaîne télévisée <i>France 2</i>. Tourné à Lyon, il évoque certains animaux familiers menacés de disparition comme la grenouille, le grand hamster... Cependant, tu n'as pas pu le voir car tu avais des cours à préparer. Trouvant le thème de <i>la disparition des animaux</i> très intéressant, tu as voulu élaborer un dossier constitué d'articles informatifs qui seront publiés dans l'encyclopédie numérique créée par votre établissement. Ces articles contiennent des textes, de douze (12) lignes chacun, qui doivent :</p> <ul style="list-style-type: none"> <li>- Expliquer les raisons de la disparition de certaines espèces sauvages.</li> </ul>

<sup>123</sup> Les compétences de compréhension et de production indissociablement présentées dans notre étude sont séparées dans les documents d'accompagnement consultés (2016).

- Préciser les dangers de la disparition des animaux.
- Proposer des solutions qui pourraient sauver les espèces menacées.

Pour ce faire, tu as collecté les documents suivants que tu dois lire pour en extraire quelques informations.

<b>Document 1</b>	<b>Document 2</b>
<p>« Les menaces qui pèsent sur la faune et sur la flore sauvages sont, à l'heure actuelle, plus sérieuses que jamais. De nombreuses espèces animales ou végétales diminuent d'une manière inquiétante et risquent d'allonger rapidement la triste liste de celles que l'homme a exterminées dans le passé. Parmi les mammifères, trente-six espèces ont disparu à tout jamais et cent vingt sont actuellement en voie de disparition avancée. Parmi les oiseaux, quatre-vingt-quatorze espèces sont éteintes et cent quatre-vingt-sept sont en danger d'extinction ».</p> <p><i>Jean Dorst, La plupart des espèces sont menacées de disparition par suite des interventions humaines, Dans <a href="http://www.monde-diplomatique.fr">www.monde-diplomatique.fr</a>, publié en 1970</i></p>	<p>« La disparition des espèces animales est un phénomène inquiétant parce qu'il modifie beaucoup de choses dans la nature, notamment la chaîne alimentaire. Il appauvrit aussi la belle diversité qui règne sur la terre. Savoir pourquoi les animaux peuvent disparaître permet de mieux les protéger.</p> <p><b>Quels sont les animaux en danger ?</b></p> <p>Les animaux, susceptibles d'être en danger et selon leur nombre actuel, sont classés en 3 catégories :</p> <ul style="list-style-type: none"> <li>- vulnérables : il faut y faire attention et surveiller leur démographie (leur nombre)</li> <li>- en danger : il faut les protéger pour que les choses n'empire pas.</li> <li>- en critique danger d'extinction : il faut tenter de préserver l'espèce en protégeant un par un chacun de ses individus. Cette dernière catégorie est la plus grave et l'animal est alors menacé à disparaître rapidement.</li> </ul> <p>(...) Heureusement des associations internationales comme WWF et des professionnels veillent à la protection de la nature et des espèces menacées ».</p> <p><i>Graine de Curieux, Pourquoi certains animaux sont-ils en voie de disparition ? Dans <a href="http://www.idkids.fr">www.idkids.fr</a>, publié le 13/9/2017</i></p>
<p style="text-align: center;"><b>Document 3</b></p> <p>« La pollution : cela joue un rôle important dans la voie de disparition de certaines espèces. La mer étant polluée, les poissons avalent des produits toxiques, qui eux-mêmes sont mangés par des phoques, et ainsi de suite. Comme tu peux le comprendre, toute la chaîne alimentaire est touchée.</p> <p>La modification de l'environnement : en construisant ses villes, l'homme a détruit des forêts entières, des lieux de nature, des espaces qui abritent beaucoup d'espèces. De même, les gaz toxiques rejetés par les usines ont un impact sur le réchauffement de la planète et donc sur les glaciers. La fonte des glaciers est terrible pour certaines espèces, comme les ours polaires.</p> <p>La pêche et la chasse : avec une population toujours plus importante sur terre, les hommes ont besoin de se nourrir. La pêche massive impacte sur la quantité de saumons, de thons ou d'autres poissons dans la mer. Ils deviennent de plus en plus rares ».</p> <p><i>Graine de Curieux, Pourquoi certains animaux sont-ils en voie de disparition ? Dans <a href="http://www.idkids.fr">www.idkids.fr</a>, publié le 13/9/2017</i></p>	<p style="text-align: center;"><b>Document 4</b></p> <p>« Le <b>tigre</b> est l'un des animaux les plus menacés dans le monde, chassé pour sa fourrure, ses yeux, ses os et même ses organes...</p> <p>Les <b>tortues</b> : (...) Les tortues confondent souvent les <b>plastiques</b> flottant dans l'océan avec de la nourriture, qui finissent par provoquer leur mort.</p> <p>L'<b>éléphant</b> : (...) En raison de la déforestation et d'une chasse incontrôlée, <b>il risque de disparaître au cours des vingt prochaines années...</b> ».</p> <p><i>Julia Perez, Les animaux en voie de disparition, Dans <a href="http://www.ompe.org">www.ompe.org</a>, publié le 2/2/2019</i></p>

Dans tes productions, il est important de mettre les verbes au présent de l'indicatif, employer les connecteurs logiques et chronologiques, investir une ponctuation appropriée et respecter la structure des textes à visée argumentative.

**Les critères de correction**, que contient la grille de correction, sont à communiquer aux apprenants.

### Situation 03

<b>Compétence évaluée</b>	
Comprendre et produire <b>des textes argumentatifs</b> en tenant compte des contraintes de la situation de communication	
<b>Épreuve</b>	
<b>Établissement</b> : ..... <b>Type</b> : ..... <b>Matière</b> : ..... <b>Niveau</b> : ..... <b>Durée</b> : .....	
<b>Évaluation</b>	
<p>La propreté des plages est un sujet qui revient chaque année en été étant donné que les vacances estivales des touristes sont parfois gâchées notamment quand ces derniers se retrouvent dans des plages sales où les déchets et les ordures sont jetés partout. Comme la situation devient alarmante, l'association <i>Pour une vie meilleure</i> appelle les bénévoles, dont tu fais partie, à réaliser des fiches explicatives, qui seront affichées dans les plages polluées, pour sensibiliser les citoyens à la nécessité de préserver leur propreté. Pour réaliser ce projet, les tâches sont réparties sur les bénévoles sollicités par cette association. Les tiennes consistent à répondre par écrit aux questions suivantes :</p> <ul style="list-style-type: none"> <li>- Quelles sont les principales causes de cette pollution des plages et de la mer ?</li> <li>- Quels sont les gestes à éviter pour préserver la propreté des plages ?</li> <li>- Quelles sont les conséquences de la pollution de la mer ?</li> </ul> <p>Ces tâches sont à réaliser en te référant aux ressources qui suivent :</p>	
<b>Document 1</b>	<b>Document 2</b>
<p>« La mer, c'est sale ! Mais pas tellement à cause des poissons qui baisent dedans, comme le prétend Renaud. Car c'est bien l'homme qui est à l'origine de la plupart des pollutions maritimes : déversements d'hydrocarbure, eaux souillées en provenance d'usines ou d'égouts, pesticides ou matières organiques issus de l'agriculture... Mais quelles sont les conséquences sur la santé et où peut-on se baigner ? (...) Une quantité d'hydrocarbures répandus en mer provient du dégazage des bateaux, rejet intentionnel des résidus de combustion de fioul lourd. <b>En Afrique, la majeure partie des plages qui auraient pu compter parmi les plus belles de l'Atlantique, sont très gravement polluées</b> du fait</p>	<p><i>La pollution marine est malheureusement aujourd'hui un fait solidement établi qui concerne l'ensemble des mers et des océans du globe.</i>  <b>Claude Alzieu</b></p> <p>« La mer est le réceptacle de toutes sortes de déchets d'origine anthropique lesquels sont en volume croissant de telle sorte que de nos jours elle n'arrive plus à « digérer » le tout. Aujourd'hui, la pollution touche la flore et la faune à tous les niveaux et par conséquent elle finit par atteindre son initiateur : l'homme. Trois phénomènes inquiètent les spécialistes : les « marées noires », les déchets de toutes sortes et l'eutrophisation des milieux côtiers.</p>



<p>de cette pratique. À cela s'ajoute que chaque année, les navires touristiques et marchands déversent des centaines de milliers de tonnes de ballast sale, de substances nocives et d'ordures. Et ce n'est pas le pire ».</p> <p><i>Ellonore Bush, La pollution côtière, Dans <a href="http://www.routard.com">www.routard.com</a>, consulté en 2020</i></p>	<p><i>Jean-Noël Salomon, La pollution de la mer (Danger pollutions !), Dans <a href="https://books.openedition.org">https://books.openedition.org</a>, publié en 2003</i></p>
<p style="text-align: center;"><b>Document 3</b></p> <p>« Chaque année dans le monde, 8 millions de tonnes de plastiques sont déversées dans les océans, dont 200 000 tonnes en Méditerranée. Ces déchets proviennent à 80 % des zones continentales où ils sont transportés par les cours d'eau ou les intempéries. D'autres sont jetés directement en mer ou sur les plages.</p> <p>(...) Ajoutons à cela une mauvaise gestion des déchets et le tourisme de masse et <b>la Méditerranée devient l'une des mers les plus polluées au monde par les plastiques</b>. En effet, le plastique représente aujourd'hui 95% des déchets sur les plages et en surface de la mer Méditerranée ».</p> <p><i>Christophe Magdelaine, La Méditerranée : l'une des mers les plus polluées au monde par les déchets plastiques, Dans <a href="http://www.notre-planete.info">www.notre-planete.info</a>, publié en 2019</i></p>	<p style="text-align: center;"><b>Document 4</b></p> <p>La <b>pollution marine</b> consiste en la présence de déchets dans les océans, ou bien résulte du rejet dans l'environnement par les activités humaines de quantités excessives de produits physiques ou chimiques toxiques, ou de déchets abandonnés par l'homme sur les sols, qui viennent polluer les fleuves et se retrouvent finalement dans les mers.</p> <p>Cette pollution arrive dans le milieu marin via le ruissellement et les cours d'eau, ou est apportée par les vents et les pluies, ou provient de produits et objets volontairement ou accidentellement rejetés à la mer. La pollution entraîne la disparition de certaines espèces. (...)</p> <p>La majeure partie de la pollution marine est issue des pollutions urbaines, industrielles et agricoles. (...)</p> <p>La pollution marine a de multiples conséquences sur la santé humaine, sur les écosystèmes et sur les services écosystémiques ».</p> <p><i>Pollution marine, Dans <a href="https://fr.wikipedia.org">https://fr.wikipedia.org</a>, consulté en 2020</i></p>
<p>Dans tes productions, il est important de mettre les verbes au présent de l'indicatif, employer les connecteurs logiques et chronologiques, investir une ponctuation appropriée et respecter la structure des textes à visée argumentative.</p> <p>Longueur des textes produits : environ douze (12) lignes.</p>	
<p><b>Les critères de correction</b>, que contient la grille de correction, sont à communiquer aux apprenants.</p>	

## Conclusion

L'APC introduite suite à la réforme du système éducatif algérien implique un changement profond des pratiques enseignantes et évaluatives. Les situations sont en effet les outils indispensables pour que des enseignements soient dispensés (situations d'apprentissage), consolidés (situations d'intégration) puis évalués (situations d'évaluation). Ces outils, que les instructions officielles préconisent, sont cependant écartés par les enseignants. Ainsi, leurs pratiques évaluatives que nous avons étudiées dans la présente recherche révèlent que les

exigences de l'APC et les instructions officielles ne sont pas respectées étant donné que les pratiques traditionnelles antérieures à cette approche sont actuellement mises en œuvre.

En évaluation du FLE, les enseignants investissent des items pour évaluer la compréhension et des consignes de rédaction pour évaluer la production, deux compétences qui ne devraient pas être évaluées séparément. Cet écart entre les prescriptions institutionnelles et les pratiques effectives -que notre analyse des épreuves collectées confirme- justifie notre proposition de dispositifs évaluatifs qui seraient plus conformes aux exigences d'une évaluation de compétences.

Ces dispositifs consistent en des situations évaluatives évaluant la maîtrise des compétences visées à tous les niveaux du cycle et en une grille de correction que nous jugeons applicable à toutes les situations. Pour les concevoir, les directives que les travaux théoriques consultés listent sont suivies : supports inédits, tâches multiples, critères indépendants... Il en est de même pour les étapes de conception qui se réfèrent aux études évoquées dans notre travail. Cependant, les références consultées ne fournissent pas de modèles de situations évaluant les compétences en langues (et en FLE particulièrement) ; d'où la formulation de propositions inédites et sujettes à de nombreuses critiques. Certes, les experts (enseignants et inspecteurs) consultés pour la validation des situations proposées ont noté certaines insuffisances auxquelles nous avons tenté de remédier. En dépit des corrections apportées, ces outils de mesure ne sont pas idéaux dans tout contexte vu que les enseignants doivent nécessairement les adapter à la particularité de leurs apprenants, de leurs niveaux, de leurs intérêts...

Contrairement aux situations qui ne sont pas inspirées de modèles préexistants, la grille est inspirée de modèles existants dans les travaux de Gérard (2009), Côté, Tardif et Munn (2011)... Nous nous contentons d'une seule grille pour évaluer les productions des apprenants de tous les niveaux du cycle : les critères sélectionnés sont, en effet, identiques à tous les niveaux et les indicateurs qui les constituent sont généraux, donc applicables à toutes les productions quel qu'en soit le type.

Tout comme les situations, cette grille est à adapter pour être applicable aux attentes des enseignants et au niveau des apprenants. Voyant que cette grille ne ressemble pas à celles antérieurement proposées dans les documents officiels (notamment que l'échelle d'appréciation, le seuil de réussite et le jugement ne sont pas familiers aux enseignants), certaines directives sont fournies pour que les évaluateurs puissent se servir de cet outil de jugement. Certes, toutes les précisions sur les situations évaluatives et sur la grille de correction

sont à communiquer aux apprenants avant de les exploiter étant donné que ces outils ne sont pas introduits dans les épreuves actuelles évaluant le FLE au cycle moyen.

Cependant, notre initiative est loin de parvenir à un guide pratique exploitable dans tout contexte : ce ne sont que des propositions qui pourraient être enrichies par les enseignants évaluateurs. Les insuffisances que les lecteurs pourraient y soulever sont dues essentiellement au recours à une seule procédure de validation : la consultation des collègues. Si cette procédure qui pourrait à première vue peu pratique est investie, c'est en raison des pratiques enseignantes actuelles qui ne dispensent pas réellement des enseignements conformes aux instructions de l'APC. Ce constat est formulé suite à une consultation des manuels où aucune situation d'apprentissage et d'évaluation conforme aux exigences ne figure.

Il est temps alors de rénover les pratiques enseignantes qui conditionnent tout le processus évaluatif vu que toute évaluation doit porter sur ce qui est enseigné. La réticence de certains enseignants, qui perpétuent les pratiques anciennes (pré-APC) au cycle moyen, ralentit considérablement la réforme du système éducatif algérien ; d'où la nécessité de les pousser à adopter les nouvelles pratiques d'enseignement à travers le changement des pratiques évaluatives.

# **CONCLUSION GENERALE**

A travers notre étude intitulée « **La validité de l'évaluation sommative du FLE au cycle moyen en Algérie : de l'étude à la conception d'un dispositif d'évaluation** », nous avons tenté de répondre à notre problématique et à toutes les questions qui en découlent. Rappelons que notre travail tentait de préciser à quel point les dispositifs actuellement mis en œuvre en évaluation sommative sont valides et de déterminer quels outils, si les dispositifs actuels présentent des problèmes de validité, les enseignants du cycle moyen doivent proposer pour que des aperçus fidèles sur les compétences effectives de leurs apprenants en FLE soient obtenus. Par cette notion de validité, nous désignons la capacité de tous les instruments (outils de mesure et outils de jugement), intégrés tout au long du processus évaluatif, à obtenir des résultats valides, reflétant le plus fidèlement possible les compétences des apprenants évalués.

Pour ce faire, nous avons opté pour une analyse qualitative et quantitative des épreuves sommatives (devoirs et compositions) introduites au cours des années scolaires 2017/2018, 2018/2019, 2019/2020 ; c'est-à-dire après la réforme de la deuxième génération qui a apporté des changements qualitatifs profonds aux programmes, contenus, manuels et pratiques mis en œuvre suite à la première phase de réforme entamée en 2003. En empruntant cette démarche mixte, les pratiques évaluatives actuellement introduites par les enseignants du cycle moyen ont été décrites, ce qui nous a permis de révéler leur degré de conformité aux pratiques nouvelles, préconisées dans les documents officiels et dans les références citées dans notre recherche.

Les résultats primaires obtenus ont révélé que les dispositifs évaluatifs actuels présentent des problèmes de validité et que de nouveaux dispositifs plus conformes aux exigences de l'APC, intégrée dans le système éducatif algérien depuis 2003, devront être conçus pour évaluer correctement les compétences effectives des apprenants en FLE.

Cette non validité est essentiellement due au non-respect de plusieurs exigences de la nouvelle approche (APC). Bien que cette approche préconise le recours à des situations inédites, complètes, motivantes...qui évaluent indissociablement compréhension et expression, les épreuves analysées les évaluent séparément, comme dans les épreuves traditionnelles, en recourant à des outils inadaptés aux nouvelles instructions. Ainsi, la compréhension est évaluée à travers des items –portant sur des supports parfois adaptés ou fabriqués -d'une validité contestée : les QCM, les questions ouvertes...alors que la production est évaluée à travers des situations (d'intégration) incomplètes, inédites...Bien que ces situations soient déclarées conformes aux exigences de la nouvelle approche et aux instructions officielles, notre étude a noté certaines insuffisances telles que :

- L'absence de supports alternant des informations utiles et inutiles ;

## Conclusion générale

- Le non-respect de la règle des 2/3 : une seule tâche est en fait proposée dans la situation unique constituant les épreuves sommatives analysées, ce qui ne permettra pas une évaluation effective des compétences des apprenants ;
- L'imprécision du but de la production, du destinataire...

Ces deux compétences ne sont pas les seules composantes évaluées étant donné que les outils grammaticaux –désignés par ressources dans notre étude- sont également mesurés à travers des items souvent décontextualisés. Cette évaluation des ressources dans une perspective sommative, qui ne doit théoriquement inclure que les ressources importantes, ne porte que sur des savoirs et savoir-faire : ce sont notamment les niveaux cognitifs inférieurs que les enseignants visent en évaluation, ce qui n'est pas conforme aux nouvelles instructions. De surcroît, les ressources évaluées se voient accordées un taux de points qui dépasse souvent les exigences : alors que le taux de points consacré aux ressources ne doit pas dépasser 25%, les enseignants continuent à leur accorder des taux dépassant parfois 50%.

Les insuffisances relevées qui affectent la validité ne sont pas exclusives aux outils de mesure : les outils de jugement -qui doivent être clairement mentionnés et communiqués aux apprenants- font défaut dans les épreuves analysées, ce qui conduit à la subjectivité des enseignants.

Ainsi, notre analyse a confirmé la première hypothèse que nous avons précédemment formulée : la non validité des épreuves sommatives actuelles du cycle moyen qui ne respectent pas les instructions officielles inspirées des exigences de la nouvelle approche. Cette hypothèse ne constitue pas la seule que notre étude a confirmée. La confrontation des BEM pré-APC aux BEM post-APC a confirmé, en effet, notre deuxième hypothèse ; autrement dit la ressemblance supposée entre les deux catégories d'épreuves. En vue de vérifier cette ressemblance entre ces épreuves, qui est la manifestation concrète de la stagnation des pratiques évaluatives, nous avons eu recours à la confrontation des BEM post-APC, supposés conformes aux exigences, aux BEM pré-APC. Pour les BEM pré-APC, intégrés avant 2007, nous notons que la période 1999-2007 a connu le recours à deux sujets pour évaluer le FLE qui était évalué en tant que première langue étrangère (LE1) ou en tant que deuxième langue étrangère (LE2) selon le statut accordé aux deux langues étrangères en présence : français et anglais. Seules les épreuves évaluant le français comme première langue étrangère –statut actuel de cette langue- sont retenues pour une comparaison légitime.

Suite à cette comparaison, une similitude flagrante entre les deux catégories de BEM était démontrée : les outils de mesure y sont en effet quasi-identiques et les outils de jugement

## Conclusion générale

sont omis. En effet, les compétences de compréhension et de production sont évaluées séparément ; les ressources linguistiques occupent toujours un volet important et le taux de points qui leur est accordé dépasse souvent le taux déterminé (25% de la note totale) ; les situations évaluant la production ne sont pas complètes et inédites ; les outils de jugement ne sont pas communiqués aux apprenants...

La conclusion de stagnation des pratiques évaluatives, précédemment formulée, est ainsi prouvée ; d'où leur incapacité à refléter fidèlement les compétences des évalués.

Nos conclusions reconnaissent ainsi que les outils de mesure et de jugement qui constituent les épreuves de FLE proposées pour évaluer les compétences des collégiens algériens ne sont pas valides étant donné que ces outils ne sont pas adaptés aux instructions de la nouvelle approche. Ce problème de validité dont souffrent les pratiques évaluatives du cycle moyen pourrait refléter celui de l'évaluation sommative dans tous les cycles de l'Éducation nationale. Nous ne pouvons néanmoins généraliser les résultats obtenus sans réserves étant donné que le corpus retenu ne pourrait donner un aperçu sur les pratiques évaluatives au cycle primaire ou au lycée.

Une telle situation rend nécessaire une réforme urgente de l'évaluation scolaire pour concevoir des dispositifs évaluatifs valides donnant des aperçus fidèles sur les compétences des apprenants algériens en FLE, qui ne serait pas la seule discipline concernée par le changement étant donné que toutes les langues enseignées (arabe et anglais) sont évaluées suivant la même démarche. Ainsi, nos suggestions émises au niveau de notre introduction générale se trouvent infirmées dans la mesure où une rénovation radicale est nécessaire.

Étant donné que la compréhension et la production sont des compétences mobilisées indissociablement dans les pratiques communicatives, les évaluer séparément ne fournit pas un aperçu fidèle sur les compétences effectives des évalués ; d'où la nécessité de recourir à des situations évaluatives, distinctes des situations d'intégration proposées dans les manuels scolaires et présentées, à tort, comme des situations évaluatives. Les items sélectionnés pour évaluer la compréhension, les ressources linguistiques et la production –actuellement introduits dans les épreuves sommatives et dans les BEM post-APC- deviennent inadaptés aux exigences de l'APC. De ce fait, les tests du CECRL -qui recourent à certains types de questions (QCM, vrai/faux...) reconnus valides- le deviennent également.

## Conclusion générale

Il devient alors nécessaire de concevoir des dispositifs valides, évaluant efficacement les compétences des apprenants algériens de FLE, d'où la nécessité de déterminer les outils auxquels les enseignants peuvent recourir. Comme la validité des dispositifs évaluatifs requiert celle des outils de mesure et de jugement les constituant, des situations et des grilles de correction qui remplissent des exigences précises sont à proposer.

S'agissant des situations, elles doivent être :

- Complexes : présentant des problèmes à résoudre, mobilisant les opérations cognitives supérieures, mobilisant les ressources enseignées...
- Signifiantes : présentant un contexte lié au vécu des apprenants, constituant un défi à surmonter...
- Authentiques : proposant des tâches réalistes, aboutissant à une production socialisable...
- Complètes : précisant le contexte, contenant des supports variés, proposant au moins trois tâches indépendantes à résoudre...
- Inédites : contenant des supports inédits, proposant de nouvelles tâches...
- Adéquates : évaluant des compétences retenues dans les programmes, véhiculant des valeurs positives...

Pour les grilles de correction, elles doivent :

- Préciser les critères de correction : ces critères doivent être pertinents pour la discipline évaluée. Ainsi, des critères comme la correction linguistique, la cohérence... qui sont pertinents en langues ne le sont pas en sciences, en mathématiques...
- Englober des critères minimaux et des critères de perfectionnement : les critères sélectionnés qui ne doivent pas être nombreux sont répartis en deux catégories. La première englobe les critères minimaux dont la maîtrise est indispensable pour que la compétence soit maîtrisée alors que la seconde est constituée de critères de perfectionnement qui ne sont pas nécessaires à la maîtrise de la compétence.
- Accorder aux critères minimaux le  $\frac{3}{4}$  de tous les critères retenus : les points accordés aux critères minimaux ne doivent pas dépasser le quart de la note totale.
- Préciser les indicateurs liés à chaque critère : leur précision est nécessaire pour que les critères soient compris par les enseignants correcteurs.
- Éclaircir l'échelle de mesure : le type de l'échelle de mesure est à préciser pour que les enseignants puissent investir correctement la grille conçue.



A ce niveau de notre conclusion, il convient de préciser que les barèmes de notation et les listes de vérification ne constituent pas des instruments adaptés à une évaluation sommative rigoureuse des productions des apprenants. D'ailleurs, les notes -que nous avons supposées inappropriées dans une perspective de compétences- se sont avérées très pratiques. Les enseignants doivent néanmoins les appuyer par des expressions phrastiques précisant si les compétences évaluées sont réellement maîtrisées par les apprenants.

Comme les références consultées ne fournissent pas de modèles de situations évaluant les compétences en langues (et en FLE particulièrement), nous avons proposé –en obéissant aux critères ci-dessus listés- un dispositif constitué de situations et d'une grille.

Les situations conçues ont été corrigées en tenant compte des commentaires formulés par les experts consultés pour les valider. En dépit des corrections apportées, ces situations ne pourraient être appliquées à tout contexte vu que les enseignants doivent les adapter à la particularité de leurs apprenants, de leurs niveaux, de leurs intérêts...

Contrairement aux situations qui ne sont pas inspirées de modèles préexistants, la grille est inspirée de modèles exposés dans les travaux de Gérard (2009), Côté, Tardif et Munn (2011)... Notons qu'une seule grille est proposée pour évaluer les productions des apprenants de tous les niveaux du cycle : les critères sélectionnés sont, en effet, identiques à tous les niveaux et les indicateurs qui les constituent sont généraux, applicables à toutes les productions à élaborer par les apprenants. Tout comme les situations, cette grille est à adapter pour être applicable aux attentes des enseignants et au niveau des apprenants. Précisons que certaines directives, jugées nécessaires, sont fournies pour que les évaluateurs puissent se servir de cet outil de jugement. Certes, toutes les précisions sur les situations évaluatives et sur la grille de correction sont à communiquer aux apprenants avant de les exploiter étant donné que ces outils ne sont pas introduits dans les épreuves actuelles évaluant le FLE au cycle moyen.

Bien que les outils de mesure (situations) et de jugement (grilles de correction) proposés soient plus conformes aux exigences de la nouvelle approche (APC) appliquée que les outils actuellement investis par les enseignants du cycle, les critiques ne pourraient pas les épargner étant donné que les acteurs de la classe de FLE pourraient les trouver inadaptés à leurs besoins et à leurs niveaux. Ces critiques seraient légitimes si nous tenons compte de la procédure de validation adoptée : la consultation des experts (inspecteurs)<sup>124</sup>. Celle-ci ne contribue pas à

---

<sup>124</sup> Les inspecteurs sollicités pour délivrer des rapports validant les situations proposées se sont contentés de donner quelques remarques que nous intégrerons dans la partie (Annexes) sans les signer. Cependant, comme les

## Conclusion générale

collecter les avis des apprenants, leurs commentaires, leurs questions...qui sont plus constructives dans la validation des épreuves évaluatives. Si cette démarche –en dépit de ses limites- est investie, c'est en raison des pratiques actuelles des enseignants -exerçant au cycle moyen- qui ne voulaient pas expérimenter ces outils extrêmement distincts de leurs pratiques habituelles ; d'où la difficulté de les proposer à leurs apprenants. Cette difficulté est significative car elle révèle que la rénovation des pratiques évaluatives doit être progressive.

Pour que les nouvelles pratiques évaluatives soient introduites, il est important alors de rénover les pratiques enseignantes qui conditionnent tout le processus évaluatif vu que toute évaluation doit porter sur ce qui est enseigné. La réticence de certains enseignants, qui perpétuent les pratiques anciennes (pré-APC) au cycle moyen, ralentit considérablement la réforme du système éducatif algérien ; d'où la nécessité de les pousser à adopter les nouvelles pratiques d'enseignement à travers le changement des pratiques évaluatives. Cette idée serait admise si nous tenons compte de la tendance qu'ont de nombreux enseignants à copier leurs pratiques enseignantes sur les pratiques évaluatives.

Ce changement ne pourrait néanmoins être entamé sans une formation préalable des enseignants, des formateurs et des inspecteurs pour les amener à intégrer pleinement les pratiques enseignantes préconisées par la nouvelle approche et les préparer à de nouvelles modalités d'évaluation ; une formation qui nécessite le recours à des experts et à des études qui en déterminent les étapes, des contenus, la durée...

De nombreuses études nationales sont, également, à mener pour déterminer si les résultats auxquels est parvenu le présent travail s'appliquent aux autres cycles (le primaire et le secondaire) ; autrement-dit si les dispositifs investis en évaluation sommative présentent les mêmes problèmes de validité et requièrent une rénovation des outils de mesure et de jugement. Cette réforme des pratiques évaluatives doit être une priorité car elle contribue à propulser la réforme des pratiques des enseignants dans les cycles qui stagnent.

---

enseignants consultés, ces remarques révèlent que les inspecteurs –qui assurent la transmission des instructions officielles aux enseignants- ne sont pas au courant de toutes les exigences de la nouvelle approche (APC).

# **BIBLIOGRAPHIE**

## Ouvrages

1. Adel, F. (2005). L'élaboration des nouveaux programmes scolaires. Dans Toualbi-Thaâlibi & Tawil (dir.), *La refonte de la pédagogie en Algérie : Défis et enjeux d'une société en mutation* (45-56). Rabat : UNESCO.
2. AERA, APA, & NCME (1999). *The Standards for Educational and Psychological Testing*. Washington : AERA Publications Sales.
3. Allal, L. (2000). Acquisition et évaluation des compétences en situation scolaire. Dans J. Dolz & E. Ollagnier (dir.), *L'énigme de la compétence en éducation* (77-94). Bruxelles : De Boeck.
4. Allal, L. & Mottier Lopez, L. (2008). Mieux comprendre le jugement professionnel en évaluation : apports et implications de l'étude genevoise. Dans *Le jugement professionnel en évaluation- Pratiques enseignantes au Québec et à Genève*. Québec : Presses de l'Université du Québec.
5. Almond, R. G., Mislevy R. J., Steinberg, L., Yan, D. & Williamson, D. (2015). *Bayesian Networks in Educational Assessment*. New York : Springer.
6. Angers, M. (2015). *Initiation pratique à la méthodologie des sciences humaines*. Alger : Casbah.
7. Antibi, A. (2003). *La constante macabre, ou comment a-t-on découragé des générations s'élèves ?* Toulouse : Math 'Adore.
8. Ardouin, T. (2013). *Ingénierie de formation : Analyser, concevoir, réaliser, évaluer* (4ème édition). France : DUNOD.
9. Aubin, A. S. (2015). *Caractéristiques de l'évaluation des apprentissages dans une faculté de génie*. Thèse de doctorat en éducation. Université de Sherbrooke.
10. Aubret, J., Gilbert, P. & Pigeyre, F. (1993). *Savoir et pouvoir : les compétences en question*. Paris : PUF.
11. Beckers, J. (2002). Comment former les enseignants à l'évaluation des compétences ? Dans L. Paquay, Gh. Carlier, L. Collès & A-M. Huynen (dir.), *L'évaluation des compétences chez l'apprenant : Pratiques, Méthodes et Fondements* (147- 153). Louvain-la Neuve : Presses universitaires de Louvain.
12. Beckers, J. (2002). *Développer et évaluer des compétences à l'école : vers plus d'efficacité et d'équité*. Bruxelles : Labor.
13. Bélair, L. M. (1993). *Profil d'évaluation*. Montréal : Les Éditions de la Chenelière inc.
14. Bélair, L. M. (1999). *L'évaluation dans l'école*. Paris : ESF.

15. Bélair, L. (2014). Evaluer ce qu'ils ont appris. Dans L. Ménard & L. St-Pierre (dir.), *Se former à la pédagogie de l'enseignement supérieur* (355-380). Montréal : ChenelièreEducation.
16. Bélair, L. M. (2015). Entre lotto-évaluation et autoévaluation : questionnements sur les logiques de formations et les démarches d'accompagnement. Dans P-F. Coen & L. M. Bélair (dir.), *Evaluation et autoévaluation- Quels espaces de formation ?* (37-54). Louvain-La-Neuve : De Boeck supérieur.
17. Bellier, S. (1999). La compétence. Dans P. Carré & P. Caspar (Eds.), *Traité des sciences et des techniques de la formation* (223-244). Paris : Dunod.
18. Bentaïfour, B. (2009). *Didactique du texte littéraire : choisir et exploiter un texte pour la classe*. ThalaEditions.
19. Bernier, J. J. & Pietrulewicz, B. (1997). *La psychométrie. Traité de mesure appliquée*. Montréal : Gaëtan Morin éditeur.
20. Bloom, B. S., Hastings, J. T & Madaus, G. (1971). *Handbook on formative and summative evaluation of student learning*. New York: Mac Graw Hill.
21. Bloom, B. S., Madaus, George F. & Hastings, J. T. (1981). *Evaluation to Improve Learning* (1st edition). New York: McGraw-Hill.
22. Bonniol, J. J. (1886). *A la recherche de la qualité : fonctionnements par objectifs et évaluation*. Ronéo : Université Aix-en-Provence.
23. Bonniol, J-J. & Vial, M. (2009). *Les modèles de l'évaluation : textes fondateurs avec commentaires* (2ème édition). Bruxelles : De Boeck.
24. Bourguet, E. (2017). Comment intégrer des éléments contextuels lors de l'évaluation individuelle d'étudiants en cours de langue étrangère ? Dans V. Roulin, A-C. Allin-Pfister & D. Berthiaume (dir.), *Comment évaluer les apprentissages dans l'enseignement supérieur professionnalisant ? Regards d'enseignants* (91-99). Louvain-La-Neuve : De Boeck supérieur.
25. Buisson, F. (1911). *Le Nouveau dictionnaire de pédagogie et d'instruction primaire*. Repéré à <http://www.inrp.fr/edition-electronique/lodel/dictionnaire-ferdinand-buisson>
26. Cardinet, J. (1983). *Evaluer les conditions d'apprentissage des élèves plutôt que leurs résultats*. Neuchâtel : I.D.R.P.
27. Cardinet, J. (1986). *Evaluation scolaire et mesure*. Bruxelles : De Boeck-Wesmael.
28. Cardinet, J. (1986). *Pour apprécier le travail des élèves*. Bruxelles : De Boeck-Wesmael.

29. Cardinet, J. (1990). Évaluation externe, interne, ou négociée ? In Institut romand de recherche et de documentation pédagogique (éd.), *Hommage à Jean Cardinet* (139-157). Neuchâtel : IRDP - Delval.
30. Cardinet, J. (1990). *Remettre le quantitatif à sa place en évaluation scolaire*. Neuchâtel : Institut romand de recherche et de documentation pédagogique.
31. Carette, V. (2009). Et si on évaluait des compétences en classe ? A la recherche du cadrage instruit. Dans L. Mottier Lopez & M. Crahay (dir.), *Evaluations en tension*. Bruxelles : De Boeck.
32. Carlier, Gh. & Parmentier, Ph. (2002). Quelles objections et quelles difficultés les enseignants ont-ils face à l'évaluation des compétences ? Dans L. Paquay, Gh. Carlier, L. Collès & A-M. Huynen (dir.), *L'évaluation des compétences chez l'apprenant : Pratiques, Méthodes et Fondements* (125- 132). Louvain-la Neuve : Presses universitaires de Louvain.
33. Conseil de l'Europe. (2001). *Cadre européen commun de référence : apprendre, enseigner, évaluer*. Strasbourg : Editions Didier.
34. Conseil de l'Europe. (2011). *Manuel pour l'élaboration et la passation de tests et d'examens de langue à utiliser en liaison avec le CECR*. Division des politiques linguistiques, ALTE.
35. Côté, R., Tardif, J. & Munn, J. (2011). *Élaboration d'une grille d'évaluation*. Atelier pédagogique à l'intention des enseignants universitaires. Groupe ECEM.
36. Covington, M. V. (1992). *Making the grade: a self-worth perspective in motivation and school reform*. New York : Cambridge University.
37. Crahay, M. (1996). *Peut-on lutter contre l'échec scolaire ?* Bruxelles : De Boeck Université.
38. Cronbach, L. J. (1971). Test validation. Dans R. Thorndike (Ed.), *Educational Measurement* (2nd ed., p. 443). Washington DC : American Council on Education.
39. Cronbach, L. J. (1980). Validity on parole: how can we go straight. Dans W. B. Schrader (dir.), *New directions for testing and measurement: measuring achievement: progress over a decade*. San Francisco, Californie : Jossey-Bass.
40. Cuq, J-P. & Gruca, I. (2008). *Cours de Didactique Du Français Langue Étrangère Et Seconde*. Grenoble : Presses Universitaires Grenoble.
41. Cuq, J-P. (2003). *Dictionnaire de didactique du français langue étrangère et seconde*. Paris : CLE international.

## Bibliographie

42. Danvers, F. (1994). *700 mots-clefs pour l'éducation : 500 ouvrages recensés (1981-1991)*. France : Presses Universitaires de Lille.
43. Dany, É. (2013). De la compétence communicative vers la compétence informationnelle, Ouverture et opportunité. Dans G. Carlier, M. De Kesel, J. L. Dufays & B. Wiame (dir.), *Progression et transversalité : Comment (mieux) articuler les apprentissages dans les disciplines scolaires ?* (133-140). Presses universitaires de Louvain.
44. De Ketele, J. M. (1979). L'enseignement supérieur au 21ème siècle (document d'orientation). Dans BRENDA (édi.), *Consultation de la région Afrique préparatoire à la conférence mondiale sur l'enseignement supérieur* (1-25). Dakar : BRENDA.
45. De Ketele, J-M., Chastrette, M., Cros, D., Mettelin, P. & Thomas, J. (1988). *Guide du formateur*. Bruxelles : De Boeck-Wesmael.
46. De Ketele, J-M. & Dufays, J-L. (2003). Vers de nouveaux modes d'évaluation des compétences. Dans L. Colles, J. L. Dufays & C. Maeder (dir.), *Enseigner le français, l'espagnol et l'italien, Les langues romanes à l'heure des compétences*. Bruxelles : Éditions De Boeck.
47. De Ketele, J-M. (2013). Evaluer les apprentissages dans la formation des professionnels de la santé. Dans J. Jouquan & F. Parent (dir.), *Penser la formation des professionnels de la santé* (285-303). Bruxelles : De Boeck.
48. De Landsheere, G. (1976). *La Formation des enseignants de demain*. Paris : Casterman.
49. De Landsheere, G. (1979). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris : Presses Universitaires de France.
50. De Landsheere, G. (1982). *La recherche expérimentale en éducation*. Lausanne : UNESCO.
51. De Landsheere, G. (2000). *Evaluation continue et examens - Précis de docimologie*. Bruxelles : Editions LABOR.
52. De Landsheere, V. & G. (1975). *Définir les objectifs de l'éducation*. Paris : Presses Universitaires de France.
53. De Landsheere, V. (1988). *Faire réussir, faire échouer : la compétence minimale et son évaluation*. Paris : PUF.
54. De Landsheere, V. (1992). *L'éducation et la formation*. Paris : PUF.
55. De Vecchi, G. (2011). *Evaluer sans dévaluer et évaluer les compétences*. Paris : Hachette Education.

56. Delory, Ch. (2002). L'évaluation des compétences dans l'enseignement fondamental. De quoi parle-t-on ? Dans L. Paquay, G. Carlier, L. Collès & A-M. Huynen (dir.), *L'évaluation des compétences chez l'apprenant. Pratiques, Méthodes et Fondements* (21-35). Belgique : Presses Universitaires de Louvain.
57. Derradji, Y. (2002). *Le français en Algérie. Lexique et dynamique des langues*. Bruxelles : Duculot.
58. Dionne, E. & Raïche, I. (2017). *Mesure et évaluation des compétences en éducation médicale- Regards actuels et prospectifs*. Québec : Presses Universitaires du Québec.
59. Dolz, J. & Schneuwly, B. (1998). *Pour un enseignement de l'oral : initiation aux genres formels de l'oral*. Paris : ESF éditeur.
60. Dolz, J. (2002). L'évaluation des compétences chez l'apprenant : bilan et questions à propos de la méthodologie. Dans L. Paquay, Gh. Carlier, L. Collès & A-M. Huynen (dir.), *L'évaluation des compétences chez l'apprenant : Pratiques, Méthodes et Fondements* (135- 146). Louvain-la Neuve : Presses universitaires de Louvain.
61. Durand, M. J. & Chouinard, R. (2006). *L'évaluation des apprentissages. De la planification de la démarche à la communication des résultats*. Montréal : HMH.
62. Ferhani-Meghraoui, F. F. (2019). L'évaluation des épreuves du baccalauréat à l'aune du Référentiel général des programmes. Dans Guendouz-Benammar (dir.), *Le baccalauréat en Algérie et l'évaluation des compétences. Cas du français langue étrangère (FLE)*, (21-47). Alger : CRASC.
63. Figari, G. (1980). *Evaluation : quel référentiel ?* Bruxelles : De Boeck.
64. Frenay, M. & Bédard, D. (2004). Des dispositifs de formation s'inscrivant dans la perspective d'un apprentissage et d'un enseignement contextualisés pour favoriser la construction de connaissances et leur transfert. Dans A. Presseau & M. Frenay (éd.), *Le transfert des apprentissages : comprendre pour mieux intervenir* (241- 268). Québec : Presses de l'Université Laval.
65. Galiana, D. (2005). *Mémento de l'évaluation : analyser et améliorer sa pratique de l'évaluation*. Dijon : educagri éditions.
66. Galisson, R. & Coste, D. (1976). *Dictionnaire de didactique des langues*. Paris : Hachette.
67. Gérard, F. M. (2006). L'évaluation des acquis des élèves dans le cadre de la réforme éducative en Algérie. Dans N. Toualbi-Thaâlibi (dir.), *Réforme de l'Éducation et Innovation Pédagogique en Algérie* (85-124). UNESCO-ONPS.



## Bibliographie

68. Gérard, F. M. (2006). L'évaluation des acquis des élèves dans le cadre de la réforme éducative en Algérie. Dans N. Toualbi-Thaâlibi (dir.), *Réforme de l'Education et Innovation Pédagogique en Algérie* (85-124). UNESCO-ONPS.
69. Gérard, F.M. (2008). La complexité d'une évaluation des compétences à travers des situations complexes : nécessités théoriques et exigences du terrain. Dans M. Ettayebi, R. Operti & P. Jonnaert (éd.), *Logique de compétences et développement curriculaire : débats, perspectives et alternative pour les systèmes éducatifs* (167-183). Paris : Le Harmattan.
70. Gérard, F. M. (2009). *Évaluer les compétences : Guide pratique*. Louvain-La-Neuve : De Boeck.
71. Gérard, F-M & Roegiers X. (1993). *Des manuels scolaires pour apprendre – Concevoir, évaluer, utiliser*. Bruxelles : De Boeck.
72. Gilliéron Giroud, P. & Ntamakiliro, L (Eds). (2010). *Reformer l'évaluation scolaire : mission impossible ?* Berne : Editions scientifiques internationales.
73. Guba, E. G. & Lincoln, Y. S. (1991). *Effective evaluation*. San Francisco: Jossey-Bass pub.
74. Guilford, J. P. & Fruchter, B. (1973). *Fundamental statics in psychology and education* (5ème édition). New York: Mc Graw Hill.
75. Gumbel, P. (2010). *On achève bien les écoliers*. Paris : Grasset.
76. Hadji, C. (2000). *L'évaluation, règles du jeu*. Paris : ESF.
77. Hadji, C. (2012). *Faut-il avoir peur de l'évaluation ?* Louvain-La-Neuve : de boeck.
78. Hadji, C. (2015). *L'évaluation à l'école- Pour la réussite de tous les élèves*. Paris : Nathan.
79. Hartog, P. J. (1918). *Examinations and their relation to culture and efficiency*. London, UK: Constable.
80. Howe, R., & Ménard, L. (1993). *Croyances et pratiques en évaluation des apprentissages*. Montréal : Collège Montmorency.
81. Hotyat, F. (1962). *Les examens- Les Moyens d'Evaluation dans l'Enseignement*. Paris : Editions Bourrelie.
82. Huver, E. & Springer, C. (2011). *L'évaluation en langues : Nouveaux enjeux et perspectives*. Paris : Editions Didier.
83. Jeffrey, D. (2013). *L'éthique dans l'évaluation scolaire*. Québec : Les Presses de l'Université Laval.

## Bibliographie

84. Jorro, A. (2000). *L'enseignant et l'évaluation- Des gestes évaluatifs en question*. Bruxelles : de boeck.
85. Kane, M. T. (1990). *An Argument-based Approach to Validation*. Iowa City: The American College Testing Program.
86. Kane, M. T. (2006). Validation. Dans R. L. Brennan (Ed.), *Educational Measurement*, (4th ed.), (17-64). Westport, CT: American Council on Education and Praeger.
87. Kelley, T. L. (1927). *Interpretation of educational measurements*. New York: Macmillan.
88. Lado, R. (1961). *Language testing: The construction and use of foreign language tests- A teacher's book*. New York: McGraw-Hill.
89. Lafortune, L. & Allal, L. (2008). *Jugement professionnel en évaluation : pratiques enseignantes au Québec et à Genève*. Québec : Presses de l'Université du Québec.
90. Langouet, G. & Porlier, J-C. (1981). *Mesure et statistique en milieu éducatif* (4ème édition). Paris : ESF éditeur.
91. Lasnier, F. (2014). *Les compétences de l'apprentissage à l'évaluation*. Montréal : Guérin universitaire.
92. Laurencelle, L. (1998). *Théories et techniques de la mesure instrumentale*. Ste-Foy : Presses de l'Université du Québec.
93. Laurier, M. D., Tousignant, R. & Morissette, D. (2005). *Les principes de la mesure et de l'évaluation des apprentissages* (3<sup>ème</sup> édition). Montréal : Gaëtan Morin.
94. Laveault, D. & Grégoire, J. (2002). *Introduction aux théories des tests en psychologie et en sciences de l'éducation* (2ème édition). Bruxelles : de boeck.
95. Laveault, D. (2007). De la régulation au réglage : étude des dispositifs d'évaluation favorisant l'autorégulation des apprentissages. Dans L. Allal & L. Mottier Lopez (Eds.), *Régulation des apprentissages en situation scolaire et en formation* (207-234). Bruxelles : De Boeck.
96. Laveault, D. (2007). Quelles compétences, Quels types de validité pour l'évaluation ? Dans L. Bélaïr, D. Laveault & C. Lebel (dir.). *Les compétences professionnelles en enseignement et leur évaluation* (25-47). Canada : Les Presses de l'Université d'Ottawa.
97. Le Boterf, G. (1994). *De la compétence : essai sur un attracteur étrange*. Paris : Editions d'Organisation.
98. Legendre, R. (2005). *Dictionnaire actuel de l'éducation* (3<sup>ème</sup> édition). Montréal : Guérin.

99. Lemenu, D. & Heinen, E. (2015). *Comment passer des compétences à l'évaluation des acquis des étudiants ? Guide méthodologique pour une approche programme dans l'enseignement supérieur*. Louvain-La-Neuve : De Boeck Supérieur.
100. Leplat, J. (2000). *L'analyse psychologique de l'activité en ergonomie : aperçu sur son évolution, ses modèles et ses méthodes*. Toulouse : Octarès.
101. Leroux, J. L. & Bélair, L. (2015). Exercer son jugement professionnel en enseignement supérieur. Dans J. L. Leroux (dir.), *Evaluer les compétences au collégial et à l'université : un guide pratique* (65-104). Montréal : Chenelière Education.
102. Leroux, J. L. & Bigras, N. (2003). *L'évaluation des compétences : une réalité accessible dans nos collègues*. Québec : Regroupement des collèges PERFORMA.
103. Leroux, J. L. & Mastracci, A. (2015). Concevoir des grilles d'évaluation à échelle descriptive. Dans J. L. Leroux (dir.), *Evaluer les compétences au collégial et à l'université : un guide pratique* (197-249). Montréal : QC Chenelière Education.
104. Leroux, J. L. (2010). *L'évaluation des compétences au collégial : un regard sur les pratiques évaluatives*. Canada : CEGEP de Saint-Hyacinthe.
105. Louis, R. (1999). *L'évaluation des apprentissages en classe : Théorie et pratique*. Québec : Editions Etudes vivantes.
106. Lussier, D. & Turner, C. (1995). *Le point sur... L'évaluation en didactique des langues*. Anjou : Centre éducatif et culturel.
107. Lussier, D. (1992). *Evaluer les apprentissages dans une approche communicative*. Paris : Hachette.
108. Médioni, M-A. (2016). *L'évaluation formative au cœur du processus d'apprentissage : Des outils pour la classe et pour la formation*. Lyon : Chronique Sociale.
109. Meirieu, P. (1989). *Apprendre...oui mais comment*. Paris. ESF.
110. Messick, S. (1989). Validity. Dans R. L. Linn (dir.), *Educational measurement* (3ème éd.). New York: American council on education et Macmillan.
111. Michel, S. & Ledru, M. (1991). *Capital compétence dans l'entreprise, une approche cognitive*. Paris : ESF.
112. Miled, M. (2005). Un cadre conceptuel pour l'élaboration d'un curriculum selon l'approche par les compétences. Dans N. Toualbi-Thaâlibi & S. Tawil (dir.), *La refonte de la pédagogie en Algérie- Défis et enjeux d'une société en mutation* (33-44). UNESCO-ONPS.

113. Ministère de l'Éducation et de la Formation de Manitoba. (2005). *Des outils pour favoriser les apprentissages*. Ouvrage de référence pour les écoles de la maternelle à la 8ème année. Bibliothèque nationale du Canada.
114. Mislevy, R. J. & Riconscente, M. M. (2005). *Evidence-Centered Assessment Design: Layers, Structures, and Terminology*. SRI.
115. Mislevy, R. J. & Riconscente, M. M. (2006). Evidence-Centered Assessment Design. Dans S. M. Downing & T. M. Haladyna, *Handbook of test development* (61-90). New Jersey : Lawrence Erlbaum Associates.
116. Mislevy, R. J., Steinberg, L. S., Almond, R. G. & Lukas, J. F. (2006). Concepts, terminology, and basic models of evidence-centered design. Dans D. M. Williamson, R. J. Mislevy & I. I. Bejar (dir.), *Automated scoring of complex tasks in computer-based testing*. Mahwah, New Jersey : Lawrence Erlbaum Associates.
117. Mongeau, P. (2008). *Réaliser son mémoire ou sa thèse*. Québec : Presses de l'Université du Québec.
118. Morissette, D. & Laurencelle, L. (1993). *Les examens de rendement scolaire*. Québec : Les Presses de l'Université Laval.
119. Mothe, J-C. (1975). *L'évaluation par les tests dans la classe de français*. Paris : Hachette.
120. Nadeau, M-A. (1988). *L'évaluation vue dans la perspective des programmes*. Dans *L'évaluation de programmes théorie et pratique* (2ème édition). Québec : Les Presses de l'Université Laval.
121. Ndinga, P. (2011). Validité du jugement professionnel des enseignants du primaire dans un contexte d'approche par compétences. Dans R. Gilles, K. Paquette-Côté & D. Magis, *Des mécanismes pour assurer la validité de l'interprétation de la mesure en éducation : L'évaluation*. Québec : Presses de l'Université du Québec.
122. Nisbet, J. (1990). *Teaching Thinking: An introduction to the research literature*. (Spotlight no. 26). Edinburgh : SCRE.
123. Noël-Jothy, F. & Sampsonis, B. (2006). *Certifications et outils d'évaluation en FLE*. Paris : Librairies Hachette et Larousse.
124. Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge, Cambridge University Press.
125. Nunnally, J.C. (1978). *Psychometric theory* (2nd edition). New York : McGraw-Hill.

126. Ouellet, A. (1981). *Processus de recherche : une approche systémique*. Québec : Presses de l'Université du Québec.
127. Ouellet, A. (1983). *L'évaluation créative : une approche systémique des valeurs*. Québec : Presses de l'université du Québec.
128. Paquay, L. (2002). *L'évaluation des compétences chez l'apprenant : pratiques, méthodes et fondements*. Presses Universitaires de Louvain.
129. Paquay, L. (dir.). (2004). *L'évaluation des enseignants : tensions et enjeux*. Paris : Le Harmattan.
130. Paquay, L., Defêche, N. & Dufays, J-L. (2002). Comment concilier évaluation formative et évaluation certificative ? Quels apports de l'apprenant ? Dans *L'évaluation des compétences chez les apprenants : pratiques, méthodes et fondements*. Presses Universitaires de Louvain.
131. Paquette-Côté, K. & Raïche, G. (2011). Analyse de l'argumentation de la validité des inférences d'évaluation dans les politiques institutionnelles d'évaluation des apprentissages des établissements d'enseignement collégial québécois. Dans G. Raïche, K. Paquette-Côté & D. Magis (dir.), *Des mécanismes pour assurer la validité de l'interprétation de la mesure en éducation, Volume 2 : L'évaluation* (93-119). Québec : Presses de l'Université du Québec.
132. Paquette-Côté, K. & Raïche, G. (2013). Moyens relevés dans les politiques institutionnelles d'évaluation des apprentissages (PIEA) pour assurer la validité des inférences en évaluation des apprentissages du collégial. Dans G. Raïche, P. Ndinga & H. Meunier (dir.), *Des mécanismes pour assurer la validité de l'interprétation de la mesure en évaluation, Volume 3 : Aspects pratiques* (7-38). Québec : Presses de l'Université du Québec.
133. Paquette-Côté, K. (2000). *Analyse de l'argumentation de la validité des inférences d'évaluation dans les politiques institutionnelles d'évaluation des apprentissages des établissements d'enseignement collégial québécois*. Mémoire de la maîtrise en éducation (inédit). Québec : Université du Québec à Montréal.
134. Parent, F. (2008). Formation, compétences et constructivisme : le référentiel d'évaluation outil de cohérence dans les programmes de formation en santé. Dans G. Baillat, J. M. De Ketele, L. Paquay & C. Thélot (dir.), *Evaluer pour former outils. Dispositifs et acteurs*. Bruxelles : De Boeck Supérieur.

## Bibliographie

135. Paristot, J. C. (1990). Le paradigme docimologique, un frein aux recherches sur l'évaluation pédagogique. Dans CH. Delorme (dir.), *L'évaluation en questions* (37-56). Paris : ESF.
136. Pelpel, P. (1986). *Se former pour enseigner*. Paris : Bordas.
137. Peretti, A., Boniface, J., Legrand, J-A. (1998). *Encyclopédie de l'évaluation en formation et en éducation : Guide Pratique*. Paris : ESF.
138. Perrenoud, Ph. (1997). *Construire des compétences dès l'école*. Paris : ESF.
139. Perrenoud, Ph. (1999). *Dix nouvelles compétences pour enseigner*. Paris : ESF.
140. Perrenoud, Ph. (1999). *L'école saisie par les compétences*. Université de Genève : faculté de psychologie et des sciences de l'éducation.
141. Piccardo E., Berchoud, M., Cignatta, T., Mentz, O., & Pamula, M. (2011). *Parcours d'évaluation, d'apprentissage et d'enseignement à travers le CECR*. Strasbourg : Editions du Conseil de l'Europe.
142. Pilliner, A.E.G. (1968). Subjective and Objective Testing. Dans A. Davies (ed.). *Language testing symposium: A psycholinguistic approach* (19-35). Oxford: Oxford University Press.
143. Popham, W. J. (1995). *Classroom assessment: What teachers need to know*. Boston: Allyn & Bacon.
144. Posthumus, K. (1947). *Levensgehell en school*. La Haye.
145. Prément, R. (1990). *La préparation d'un cours*. Québec : Ecole Polytechnique de Montréal.
146. Prément, R., Bernard, H. & Kozanitis, A. (2009). *Enseigner à l'université dans une approche-programme : Guide à l'intention des nouveaux professeurs et chargés de cours*. Montréal : Presses internationales Polytechnique.
147. Rabardel, P., Carlin, N., Chesnais, M., Lang, N. & Pascal, M. (1998). *Ergonomie : concepts et méthodes*. Toulouse : Octarès.
148. Raïche, G. Ndinga, P. & Meunier, H. (2013). Pour assurer la validité de l'interprétation de la mesure en éducation-Aspects pratiques. Dans G. Raïche, P. Ndinga & H. Meunier (dir.), *Des mécanismes pour assurer la validité de l'interprétation de la mesure en évaluation, Volume 3 : Aspects pratiques* (1-5). Québec : Presses de l'Université du Québec.
149. Reuchlin, M. (1962). *Les méthodes quantitatives en psychologie*. Paris : PUF.
150. Reuter, Y. (éd.). (2011). *Dictionnaire des concepts fondamentaux des didactiques*. Batna : El Midad Editions.

## Bibliographie

151. Riba, P. & Mègre, B. (2014). *Démarche qualité et évaluation en langues*. Paris : Hachette Français Langue Etrangère.
152. Ricœur, P. (1995). *Le Juste*. Paris : Esprit.
153. Robert, J. P. (2008). *Dictionnaire pratique de didactique du FLE* (2ème édition revue et augmentée). Paris : OPHRYS.
154. Roegiers, X. (2000). *Une pédagogie de l'intégration*. Bruxelles : De Boeck Université.
155. Roegiers, X. (2003). *Des situations pour intégrer les acquis*. Bruxelles : De Boeck.
156. Roegiers, X. (2004). *L'école et l'évaluation*. Bruxelles : De Boeck université.
157. Roegiers, X. (2005). L'évaluation selon la pédagogie de l'intégration : Est-il possible d'évaluer les compétences des élèves ? Dans N. Toualbi-Thaâlibi & S. Tawil (dir.), *La refonte de la pédagogie en Algérie : Défis et enjeux d'une société en mutation* (107-124). Rabat : UNESCO-ONPS.
158. Roegiers, X. (2006). L'APC dans le système éducatif algérien. Dans N. Toualbi-Thaâlibi & S. Tawil, *Réforme de l'éducation et innovation pédagogique en Algérie* (51-83). UNESCO-ONPS.
159. Saucin, J. (2018). *Méthodologie de la recherche*. Master en management d'évènements. Bruxelles : CEICS.
160. Scallon, G. (1988). *L'évaluation formative des apprentissages- La réflexion*. Québec : Les Presses de l'Université Laval.
161. Scallon, G. (2000). *L'évaluation formative*. Saint-Laurent : Éditions du Renouveau Pédagogique.
162. Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. St-Laurent, QC : ERPI.
163. Scriven, M. (1967). The methodology of evaluation. Dans R. W. Tyler, R. M. Gagné & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (39-83). Chicago, IL: Rand McNally.
164. Seghouani, A. (2005). Suivi et évaluation de la mise en œuvre de la réforme du système éducatif : vers un dispositif de pilotage. Dans N. Toualbi-Thaâlibi & S. Tawil (dir.), *La refonte de la pédagogie en Algérie : Défis et enjeux d'une société en mutation* (67-78). Rabat : UNESCO-ONPS.
165. Starch, D. (1916). *Educational measurements*. New York: Macmillan.

## Bibliographie

166. Strittmatter, A. (2007). Surveillance scolaire : quelles relations entre évaluations internes et externes ? Dans M. Behrens (Ed.), *La qualité en éducation, pour réfléchir à la formation de demain* (157-181). Québec : Presses de l'Université du Québec.
167. Stufflebeam, D. L. & al. (1980). *L'évaluation en éducation et la prise de décision*. Québec : Édition NHP.
168. Taba, H. (1962). *Curriculum development*. New York: Harcourt, Brace and World.
169. Tagliante, Ch. (2005). *L'évaluation et le cadre européen commun*. Paris : CLE international.
170. Tagliante, Ch. (2006). *La classe de langue*. Paris : CLE international.
171. Tardieu, C. (2008). *La didactique des langues en 4 mots-clés : Communication, Culture, Méthodologie, Evaluation*. Paris : Ellipses.
172. Tardif, J. (2006). *L'évaluation des compétences : documenter le parcours de développement*. Montréal : ChenelièreEducation.
173. Tawil, S. (2005). Introduction aux enjeux et aux défis de la refonte pédagogique en Algérie. Dans Toulbi-Thaâlibi & Tawil (dir.), *La refonte de la pédagogie en Algérie : Défis et enjeux d'une société en mutation* (33-44). Rabat : UNESCO.
174. Thorndike, R. L. & Hagen, E. P. (1961). *Measurement and evaluation in psychology and education* (2nd edition). New York : Wiley.
175. Toulmin, S. E. (1958). *The Uses of argument*. Cambridge, Royaume-Uni : Cambridge Universitypress.
176. Tripet Lièvre, C. (2017). Comment dépasser l'intuition pragmatique en évaluation, inhérente au domaine musical, mais problématique dans le cadre d'une formation à l'enseignement ? Dans V. Roulin, A-C, Allin-Pfister & D. Berthiaume (dir.), *Comment évaluer les apprentissages dans l'enseignement supérieur professionnalisant ? Regards d'enseignants*. Louvain La Neuve : De Boeck Supérieur.
177. Veltcheff, C. Hilton, S. (2003). *L'évaluation en FLE*. Paris : Hachette français langue étrangère.
178. Viallet, F. & Maisonneuve, J. (1981). *80 fiches pour l'évaluation de la formation et de l'enseignement*. Paris : Editions d'Organisation.
179. Vianin, P. (2009). *L'aide stratégique aux élèves en difficulté scolaire : Comment donner à l'élève les clés de sa réussite*. Bruxelles : De Boeck.
180. Vidaillet, B. (2013). *Évaluez-moi ! Évaluation au travail : les ressorts d'une fascination*. Paris : Seuil.



## Bibliographie

181. Weiss, J. (1998). *Une évaluation informative pour la régulation des apprentissages et des formations*. Neuchâtel : IRDP.
182. Wiggins, G. (1993). *Assessing Student Performance: Exploring the Purpose and Limits of Testing*. San Francisco: Jossey Bass publishers.
183. Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco : Jossey-Bass.

## Articles

1. Aït Amar Meziane, O. (2014). De la pédagogie par objectifs à l'approche par compétences : migration de la notion de compétence. *Synergies Chine*, (09), 143-153.
2. Ames, C. (1992). Classrooms: goals, structures, and student motivation. *Journal of Educational Psychology*, (84), 261-271.
3. Ammouden, M. (2018). L'approche par les compétences en Algérie : de la théorie à la pratique. *Multilinguales*, 6 (2), 117-147.
4. André, N, Loye, N. & Laurencelle, L. (2015). La validité psychométrique : un regard global sur le concept centenaire, sa genèse, ses avatars. *Mesure et évaluation en éducation*, 37 (3), 125-148.
5. Angoff, W. H. (1988). Validity: An evolving concept. Dans H. Wainer & H. I. Braun (Eds.), *Test validity* (19-32). Hillsdale, NJ: Routledge.
6. Ardouin, T. (2007). Le consultant en éducation et formation : des compétences réelles pour un métier hybride. *Les dossiers des Sciences de l'Éducation*, (18), 73-83.
7. Aylwin, U. (1995). Apologie de l'évaluation formative. *Pédagogie collégiale*, 8 (3), 24-32.
8. Baala-Boudebia, A. (2012). Langue et identité. La place du français et de l'anglais dans le conflit sociolinguistique algérien : Représentations d'enseignants de français du sud algérien. *Synergies Royaume-Uni et Irlande*, (5), 265-277.
9. Barbier, J-M. (1983). Pour une histoire et une sociologie des pratiques d'évaluation en formation. *Revue française de pédagogie*, (63), 47-60.
10. Beaume, E. (1988). Comment évaluer la lecture en primaire ? *Communication et Langages*, (77), 32-42.
11. Bélec, C. (2017). Pourquoi évaluer ? *Pédagogie collégiale*, 30 (4), 10-16.
12. Belhoste, B. (2002). L'examen : une institution sociale. *Histoire de l'éducation*, (94), 5-16.
13. Bernyer, G. & Fraisse, P. (1952). Analyse factorielle. *L'année psychologique*, 52 (2), 647-651.

14. Borsboom, D., Mellenbergh, G. J. & Van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, 111(4), 1061–1071.
15. Bouamama, M. & Boubakour, S. (2018). La réforme algérienne de la deuxième génération : défis et contraintes liés au manque de dispositifs d'évaluation de compétences. *Revue des sciences humaines*, (9), 129-147.
16. Bouchard, R. (1985). Le texte, de phrase en phrase. *Le français dans le monde*, 192, 65-71.
17. Bourque, J., Poulin, N. & Cleaver, A. (1918). Évaluation de l'utilisation et de la présentation des résultats d'analyses factorielles et d'analyses en composantes principales en éducation. *Revue des sciences de l'éducation*, 32 (2).
18. Buckingham, B. R. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, 12 (2), 71-75.
19. Buckingham, B. R., McCall, W. A., Otis, A. S., Rugg, H. O., Trabue, M. R. & Courtis, S. A. (1921). Report of the standardization committee. *Journal of Educational Research*, 4 (1), 78-80.
20. Campbell, J. T. (1962). Assessments of higher level personal: Background and scope of the research. *Personnel Psychology*, 15, 57-62.
21. Campbell, J. T., Otis, J. L., Liske, R. E. & Prien, E. P. (1962). Assessments of higher level personal: Validity of the overall assessment process. *Personnel Psychology*, 15, 63-74.
22. Campbell, N. R. (1938). Symposium: Measurement and its importance for philosophy. *Aristotelian Society Supplementary*, 17, 121-150.
23. Cardinet, J. (1975). L'élargissement de l'évaluation. Neuchâtel : IRDP. Repris (1979). *Education et Recherche*, 1 (1), 15-34.
24. Cibois, Ph. (2006). *Principe de l'analyse factorielle*. Université de Versailles. Dans <http://cibois.pagesperso-orange.fr/PrincipeAnalyseFactorielle.pdf>
25. Coulet, J-C. (2016). Compétence, compétences transversales et compétences clés : peut-on sortir de l'impasse ? *Éducation et socialisation* [En ligne], 41 | 2016. Consulté le 12 janvier 2020.
26. Crahay, M. (2006). Dangers, incertitudes et incomplétude de la logique de la compétence en éducation. *Revue française de pédagogie*, 154, 97-110
27. Cubilo, J. (2014). Argument-Based Validity in Classroom and Program Contexts: Applications and Considerations. *Shiken Research Bulletin*, 18 (1).

28. Cureton, E. E. (1951). Validity. Dans E. F. Lindquist (Ed.), *Educational measurement* (621–694). American Council on Education.
29. D'amour, C. (1996). Changer l'évaluation des apprentissages. *Pédagogie collégiale*, 9 (4), 11-15.
30. De Ketele, J. M. & Gérard, F. M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et Éducation en Évaluation*, 28 (3), 1-26.
31. De Ketele, J. M. (1996). L'évaluation des acquis scolaires : quoi ? pourquoi ? pour quoi ? *Revue tunisienne des Sciences de l'Éducation*, 23, 17-36.
32. De Ketele, J. M. (2010). La recherche scientifique en éducation : Quels critères de qualité ? *Education Sciences And Society*, 1, 21-39.
33. De Ketele, J-C., Gérard, F. & Roegiers, X. (1997). L'évaluation et l'observation scolaires- deux démarches complémentaires. *Pratiques de l'évaluation, Education*, (12), 33-37.
34. De Ketele, J-M. (1989). L'évaluation de la productivité des institutions de l'éducation. Cahiers de la Fondation universitaire. *Université et société, le rendement de l'enseignement universitaire*, 3, 73-83.
35. De Ketele, J-M. (2006). La recherche en évaluation : propos synthétiques et prospectifs. *Mesure et évaluation en éducation*, 29 (1), 99-118.
36. De Pietro, J-F. & Roth, M. (2017). A propos de la validité « didactique » d'une évaluation. *Evaluer : journal international de recherche en éducation et en formation*, 3 (3), 31-50.
37. Delory. Ch. (1991). Quelle taxonomie des compétences scolaires utiliser dans une banque d'item ? *Pédagogies*, (5), 71-100.
38. Demeuse, M. & Henry, G. (2004). Validité des mesures. Dans M. Demeuse (dir.), *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation* (163-172). Notes de cours.
39. Durand, M.J. (2008). La démarche d'évaluation dans une approche basée sur le jugement professionnel. *Vie pédagogique*, (148), 63-67.
40. Ebel, R. L. (1978). The 1978 Annual Meeting Presidential Debate: The Case For Norm-Referenced Measurements. *Educational Researcher*, 7, 3-5.
41. El Mistari, H. (2013). « L'enseignement/apprentissage du FLE en Algérie : une nouvelle méthodologie pour quelles pratiques de classe ? », *Synergies Algérie*, (18), 39-51.

42. Figari, G. (2006). L'activité évaluative entre cognition et réponse sociale : nouveaux défis pour les évaluateurs. *Mesure et évaluation en éducation*, 29 (1), 5-18.
43. Fortin, F. (1994). Propriétés métrologiques des instruments de mesure (fidélité – validité). *Recherche en soins infirmiers*, 39, 58-62.
44. Franckel, J-J. (2006). Situation, contexte et valeur référentielle. *Pratiques*, 129/130, 51-70.
45. Garcia-Debanc, C. (1999). Évaluer l'oral. *Pratiques*, (103-104), 193-212.
46. Gérard, F. M. (2002). *L'indispensable subjectivité de l'évaluation*. Antipodes, (156), 26-34.
47. Gérard, F. M. (2005). L'évaluation des compétences à travers des situations complexes. *Actes du Colloque de l'Admee-Europe*, IUFM Champagne-Ardenne, Reims, 24-26 octobre 2005.
48. Gérard, F. M. (2013). L'évaluation au service de la régulation des apprentissages : enjeux, nécessités et difficultés. *Revue française de linguistique appliquée*, (18), 75-92.
49. Guilford, J. P. (1946). New standards for test evaluation. *Educational and Psychological Measurement*, 6, 427-439.
50. Hadji, C. (1990). L'apprentissage assisté par l'évaluation : mythe ou réalité. *Cahiers pédagogiques*, (281), 20-23.
51. Hassani, Z. (2013). « La réforme du système éducatif en Algérie : quels changements dans les pratiques des enseignants ? », *Insaniyat (L'École : enjeux institutionnels et sociaux)*, (60-61), 11-27.
52. Hogan, R. T., De Fruyt, F. & Rolland, J-P. (2006). Validité et intérêt des méthodes d'évaluation de la personnalité à des fins de sélection : une perspective de psychologie appliquée aux problématiques des entreprises. *Psychologie Française*, 51(3), 245-264.
53. Kahn, S. & Rey, B. (2016). La notion de compétence : une approche épistémologique. *Education et francophonie*, 44 (2), 4-18.
54. Lenoir, Y. & Sauve, L. (1998). De l'interdisciplinarité scolaire à l'interdisciplinarité dans la formation à l'enseignement : un état de la question. *Revue française de pédagogie*, 24, (121-135).
55. Leplat, J. & Hoc, J.M. (1983). Tâche et activité dans l'analyse psychologique des situations. *Cahiers de psychologie cognitive*, 3 (1), 49-63.
56. Magherbi, H. ; Rocher, T. ; Gyselinck, V. ; Trosseille, B. & Tardieu, H. (2009). Evaluation de la compréhension de l'écrit chez l'adulte. *Economie et Statistique*, (424-425), 63-86.

57. Merle, P. (2012). L'évaluation par les notes : quelle fiabilité et quelles réformes ? *La découverte, Regards croisés sur l'économie*, 12, 218-230.
58. Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23 (2), 13-23.
59. Mislevy, R. & Haertel, G. (2006). Implications of evidence-centered design for educational testing. *Educational measurement : issues and practice*, 25 (4), 6-20.
60. Mottier Lopez, L. & Allal, L. (2008). Le jugement professionnel en évaluation : un acte cognitif et une pratique sociale située. *Revue suisse des sciences de l'éducation*, (3), 465-482.
61. Mottier Lopez, L. & Laveault, D. (2008). L'évaluation des apprentissages en contexte scolaire : développements, enjeux et controverses. *Mesure et évaluation en éducation*, 31 (3), 5-34.
62. Mottier Lopez, L. (2013). De la mesure à l'évaluation collaborative en évaluation. *Revue française d'administration publique*, (148), 939-952.
63. Newton, P. E. (2012). Clarifying the consensus definition of validity. *Measurement: Interdisciplinary Research and Perspectives*, 10 (1-2), 1-29.
64. Nonnon, E. (1998). L'apprentissage des conduites de questionnement : situations et tâches langagières. *Repères. Recherche en didactique des langues étrangères*, 17, 55-85.
65. Perrenoud, Ph. (2001). Les trois fonctions de l'évaluation dans une scolarité organisée en cycles. *Éducateur* (2), 19-25.
66. Petitjean, B. (1984). Formes et fonctions des différents types d'évaluation. *Pratiques*, 44, 5-20.
67. Petitjean, B. (1984). L'évaluation. *Pratiques*, 44, 3-4.
68. Piron, S. (2013). Contexte, situation, conjoncture. *Des contextes en histoire*. Centre de recherches historiques, 27-65.
69. Popham, W. J. (1978). The 1978 Annual Meeting Presidential Debate: The Case For Criterion-Referenced Measurements. *Educational Researcher*, 7, 6-10.
70. Rey, O. (2012). Le défi de l'évaluation des compétences. *Dossier d'actualité Veille et analyses*, 76, 1-18.
71. Richer, J-J. (2011). Recherche-action et didactique du FLE. *Synergies Chine*. (6), 47-58.
72. Roegiers, X. (2006). La pédagogie de l'intégration en bref. Rabat.

73. Romainville, M. (2011). Objectivité versus subjectivité dans l'évaluation des acquis des étudiants. *Revue Internationale de Pédagogie de l'Enseignement Supérieur* [en ligne], 27 (2). Repéré à <http://journals.openedition.org/ripes/499>
74. Ruzibiza, A. D. & De Ketele, J-M. (2007). Apprentissage par intégration des compétences de base et évaluation critériée du savoir-résumer en français langue étrangère : quels effets ? *Porta Linguarum*, (07), 13-30.
75. Shepard, L. A. (1993). Evaluating test validity. *Review of research in Education*, 19, 405-450.
76. Shepard, L.A. (2006). Classroom Assessment. Dans R. L. Brennan (Eds.), *Educational Measurement* (623-646). Westport: Praeger Publishers.
77. Smith, S. & Roodt, G. (2003). An evaluation of response scale formats of the Culture Assessment Instrument. *SA Journal of Human Resource Management*, 1(2), 60-75.
78. Soderstrom, R. & Bjork, R. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10 (2), 176-199.
79. St-Pierre, L. (2004). L'habileté d'auto-évaluation : pourquoi et comment la développer ? *Pédagogie collégiale*, 18 (1), 33-38.
80. Terman, L. M. (1915). The mental hygiene of exceptional children. *Pedagogical Seminary*, 22, 529-537.
81. Thorndike, E. L. (1916). Tests of esthetic appreciation. *Journal of Educational Psychology*, 7 (9), 509-522.
82. Tyler R. W. (1942). General Statement on Evaluation. *The Journal of Educational Research*, 35, 492-501.
83. Tyler, R. M. (1986). Changing concepts of educational evaluation. *International Journal of Educational Research*, 10, 11-35.
84. Vial, M. (2001). Recherche clinique et intervention dans les organisations : coïncidences et différences. *En Question*, 37, 6-20.
85. Viau, R. (2002). L'évaluation source de motivation ou de démotivation ? *Québec français*, (127), 77-79.
86. Voss, T., Kunter, M. and Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103, 952-969.
87. Tardif, J. (1993). *L'évaluation dans le paradigme constructiviste*. Texte tiré de L'évaluation des apprentissages. Réflexions. Nouvelles tendances et formation. Université de Sherbrooke.

### Thèses et mémoires

1. Chloé, E. (2015). *L'évaluation, entre objectivité et subjectivité : regard d'enseignants*. Mémoire de bachelor : Haute école pédagogique – BEJUNE. Repéré à <http://doc.rero.ch/record/258867>
2. Gautier, J. (2005). *Développement d'un outil d'aide à la décision environnementale durable pour une grande ville*. Mémoire de maîtrise ès sciences appliquées (génie civil). Montréal, Québec : Ecole polytechnique de Montréal.
3. Hébert, M.-H. (2013). *Validation d'une épreuve pour rendre compte du niveau de développement des compétences du programme de mathématique pour l'enseignement primaire* (thèse de doctorat non publiée). Québec, Canada : Université Laval.
4. Murillo, A. (2009). *Quels choix des enseignants quant au niveau de difficulté des tâches prescrites à leur classe ? En lecture, au Cours Préparatoire*. Thèse de doctorat. Université de Toulouse.
5. Roy, M. (2017). *Les dispositifs d'évaluation certificative mis en œuvre par des enseignants pour évaluer les attitudes dans un programme d'études collégiales : étude de cas*. Université du Québec à Montréal.

### Documents

1. Brousseau, G. (2004). *Tâche, situation, activité*. Un texte en exclusivité pour la Société Suisse pour la Recherche en Didactique des Mathématiques.
2. Carrefour de la réussite au collégial. (2004). *L'évaluation des apprentissages*. Trousse 8.
3. Commission nationale des programmes : Groupe disciplinaire de français. (2010). *Programme de français 2<sup>ème</sup> année moyenne*. Office National des Publications Scolaires.
4. Commission nationale des programmes : Groupe disciplinaire de français. (2010). *Document s'accompagnement du programme de français 1<sup>ère</sup> année moyenne*. Office National des Publications Scolaires.
5. Commission nationale des programmes : Groupe disciplinaire de français. (2010). *Programme de français 1<sup>ère</sup> année moyenne*. Office National des Publications Scolaires.
6. Commission nationale des programmes : Groupe disciplinaire de français. (2010). *Programme de français 3<sup>ème</sup> AM*. Office National des Publications Scolaires.
7. Commission nationale des programmes : Groupe disciplinaire de français. (2011-2012). *Document d'accompagnement au programmes. Français 2<sup>ème</sup> année moyenne*. Office National des Publications Scolaires.

## Bibliographie

8. Commission nationale des programmes : Groupe disciplinaire de français. (2012-2013). *Programme de français 3<sup>ème</sup> AM*. Office National des Publications Scolaires.
9. Commission nationales des programmes : Groupe disciplinaire de français. (2012-2013). *Document d'accompagnement au programme- Français 3<sup>ème</sup> AM*. Office Nationale des Publications Scolaires.
10. Commission nationale des programmes : Groupe disciplinaire de français. (2016). *Document d'accompagnement du programme de français- Cycle moyen*. Office National des Publications Scolaires.
11. Communauté française de Belgique. (1997). *Décret définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre*. Bruxelles : MERF.
12. Conseil supérieur de l'éducation au Québec. (1982-1983). *Rapport*. Québec.
13. *Guide des correcteurs/examineurs du CEFP 1, niveau A2*.
14. Lafortune, L. (2008). *L'éthique et le jugement professionnel dans le référentiel de compétences professionnelles pour l'accompagnement d'un changement prescrit*. Communication présentée à l'Association pour le développement des méthodes d'évaluation en éducation (ADMEE).
15. Mislevy, R. & Haertel, G. (2006). *Implications of evidence-centered design for educational testing*. Draft PADI technical report 17. Menlo Park, Californie : SRI International.
16. Mislevy, R. J., Almond R. G. & Lukas, J. F. (2003). *A Brief Introduction to Evidence-centered Design*. Research report. ETS Educational Testing Service.
17. Mislevy, R. J., Almond, R. G. & Lukas, J. F. (2004). *A brief introduction to evidence-centered design*. CSE Report 632. Los Angeles, Californie: The national center for research on evaluation, standards and student testing (CRESST), Center for the study of evaluation (CSE), University of California.
18. *Programme de mathématiques et de sciences physiques et chimiques pour les baccalauréats professionnels*. (2009). BO spécial N 2.
19. Voegeli-Mantovani, U. (2000). *Pour une évaluation plus formative et moins sélective : le développement de l'évaluation scolaire en Suisse*. Version abrégée du rapport de tendance du CSRE, no 3. Neuchâtel : IRDP. (00.6). 50 p.





# **ANNEXES**

**Annexe 1 : La grille d'analyse**

<b>Précisions sur l'épreuve<sup>125</sup></b>	<b>Type</b>	<b>Niveau</b>	<b>Trimestre</b>	<b>Année</b>	<b>Durée</b>	
<b>Compétences évaluées<sup>126</sup></b>	L'épreuve évalue			La compréhension		
				La production		
				Les ressources		
	La compréhension et la production sont évaluées				Séparément	
					En parallèle	
	La compréhension est évaluée à travers <sup>127</sup>				Des items	
					Des situations	
	Les ressources sont évaluées à travers				Des items	
					Des situations	
	La production est évaluée à travers				Des items	
					Des situations	
	<b>Support(s)</b>	Le nombre de supports retenus est			Un	
					Deux	
					Trois	
Ils sont proposés dans la partie					Compréhension	

<sup>125</sup> Les précisions sur l'épreuve doivent être inscrites dans les colonnes appropriées.

<sup>126</sup> Pour chacun des indicateurs constituant les rubriques de la grille, une croix (X) est à mettre devant l'élément approprié.

<sup>127</sup> Cet indicateur et les deux qui le suivent sont liés aux épreuves évaluant séparément la compréhension et la production.

Annexes

		Production	
Ils consistent en		Un texte	
		Un texte illustré	
		Une illustration	
		Un schéma	
		Autres <sup>128</sup>	
Les thèmes de la compréhension et de la production sont		Identiques	
		Liés au même domaine	
		Différents	
Ces supports existent dans le manuel scolaire		Oui	
		Non	
Le support textuel est <sup>129</sup>		Authentique	
		Adapté	
		Fabriqué	
Sa source est		Omise	
		Incomplète	
		Complète	
Sa longueur est		Moins de 15 lignes	
		De 10 à 15 lignes	
		Plus de 15 lignes	

<sup>128</sup> Si l'activité ne correspond pas aux variables proposés, son type est à insérer dans la case (Autres).

<sup>129</sup> Cet indicateur et les deux qui le suivent sont liés aux supports textuels.

<b>Outils de mesure</b>	<b>Compréhension</b>	
	Le nombre de question évaluant la compréhension est	
Les types des activités qui l'évaluent sont	QCM	
	Vrai/Faux	
	Question ouverte	
	Question fermée	
	Question orientée	
	Cases à cocher	
	Appariement	
	Classement	
Les niveaux de compréhension visés sont	Texte à trous	
	Compréhension littérale	
	Compréhension inférentielle	
	Compréhension critique	
Des questions interdépendantes existent	Compréhension créative	
	Oui	
	Non	
<b>Ressources</b>		
Le nombre de question évaluant les ressources est		
Les types des activités les évaluant sont	QCM	
	Vrai/Faux	
	Question ouverte	

		Question fermée	
		Question orientée	
		Cases à cocher	
		Classement	
		Appariement	
		Texte à trous	
	Les niveaux cognitifs des activités sont	Connaissance	
		Compréhension	
		Application	
		Analyse	
		Synthèse	
		Evaluation	
	Des questions interdépendantes existent	Oui	
		Non	
	<b>Production</b>		
	Le type des activités évaluant la production est	Rédaction	
		Remise en ordre	
		Résumé	
		Compte-rendu	
		Texte à compléter	
Réécriture			
Texte à trous			
Aucun			

Annexes

	Cette tâche proposée est	Inédite	
		Déjà traitée	
	La situation expose le contexte	Oui	
		Non	
	Le contexte de la situation précise	L'environnement	
		Le(s) destinataire(s)	
		La fonction de la production	
	Des supports sont proposés	Oui	
		Non	
	Les informations que les supports fournissent sont utiles	Oui	
		Non	
	La situation évaluative précise les instructions de réalisation	Oui	
		Non	
	Ces instructions de réalisation concernent	Les acquis à mobiliser	
		La procédure de réalisation	
		Les données techniques	
	La situation comporte trois tâches (Règle 2/3)	Oui	
		Non	
		Oui	

	Les tâches sont indépendantes <sup>130</sup>	Non	
<b>Outils de jugement</b>	L'outil de jugement investi est	Le barème de notation	
		La grille critériée	
		La liste de vérification	
	La répartition des notes est indiquée pour les activités évaluant <sup>131</sup>	La compréhension	
		La production	
		Les ressources	
	Le taux des points accordés aux ressources est	-25 %	
		+ 25 %	
		+ 50 %	
	Les critères de correction sont <sup>132</sup>	Mentionnés	
		Non-mentionnés	
	Les indicateurs relevant de chaque critère sont	Mentionnés	
		Non-mentionnés	
	Les critères retenus sont indépendants	Oui	
		Non	
Les critères minimaux sont	Précisés		
	Non-précisés		
Les critères de perfectionnement sont	Précisés		

<sup>130</sup> Cet indicateur est lié aux situations qui contiennent au moins trois tâches.

<sup>131</sup> Le deuxième et le troisième indicateurs de cette catégorie se rapportent au barème de notation.

<sup>132</sup> A partir de cet indicateur, les éléments de cette catégorie concernent un outil de jugement particulier : la grille critériée.



Annexes

		Non-précisés	
	Les deux types de critères sont	Très pertinents	
		Pertinents	
		Peu pertinents	
	Les critères de perfectionnement constituent	moins du ¼ des critères	
		Un ¼ des critères	
		Plus que le ¼ des critères	
	L'échelle de mesure est	Précisée	
		Non-précisée	

**Annexe 2 : La liste de vérification (Roegiers, 2003 ; Gérard, 2009)**

Critères	Indicateurs	Oui	Non
<i>La situation est complexe si elle...</i>	Présente un problème à résoudre.		
	Investit plusieurs ressources.		
	Intègre des ressources antérieurement enseignées.		
	Mobilise les opérations cognitives de haut niveau.		
<i>La situation est signifiante si elle...</i>	S'adresse directement à l'apprenant.		
	Implique son engagement.		
	Constitue un défi à surmonter.		
	Présente un contexte lié au vécu des apprenants.		
	Assure le transfert des apprentissages, les rendant ainsi significatifs.		
<i>La situation est authentique si elle...</i>	Traite un sujet lié à la vie réelle.		
	Présente un contexte vraisemblable.		
	Propose des tâches réalistes.		
	Vise une production socialisable.		
	Confronte l'apprenant à des contraintes semblables à celle des situations réelles.		
	Mentionne clairement le produit attendu.		
	Aboutit à une production concrète et évaluable.		
	Explicite le but de la production attendue.		

<i>La situation est complète si elle...</i>	Eclaircit le contexte de la réalisation.		
	Contient plusieurs supports variés.		
	Propose des supports contenant des informations utiles et inutiles.		
	Propose un nombre de supports identique à celui de la famille de situations.		
	Sélectionne des supports relevant du thème et du domaine de la famille de situations.		
	Contient des consignes claires et précises.		
	Contient des tâches à exécuter par l'apprenant.		
	Propose des tâches liées aux supports retenus.		
	Propose des tâches qui aboutissent à une production du type textuel étudié.		
	Exige l'exécution de trois tâches au moins.		
	Propose des tâches indépendantes.		
	Précise certaines conditions de réalisation (la durée).		
	Explicite quelques instructions de réalisation (longueur du texte).		
	Précise les critères de correction.		
	Sélectionne des critères pertinents.		
	Englobe des critères minimaux et des critères de perfectionnement.		
	Accorde aux critères minimaux le $\frac{3}{4}$ de tous les critères retenus.		
	Précise les indicateurs liés à chaque critère.		
Éclaircit l'échelle de mesure.			
	Englobe des documents inédits.		
	Propose des tâches nouvelles.		

Annexes

<i>La situation est inédite si elle...</i>	Requiert une articulation nouvelle des ressources apprises pour exécuter les tâches.		
<i>La situation est adéquate si elle...</i>	Évalue la compétence visée dans le programme.		
	Véhicule des valeurs positives.		
	Investit un vocabulaire précis et clair.		
	Est adaptée au niveau des apprenants.		

## Résumé

Notre recherche, centrée sur la validité de l'évaluation sommative du FLE en Algérie, tente d'étudier la capacité des dispositifs évaluatifs actuellement introduits par les enseignants du cycle moyen à révéler les compétences réelles de leurs apprenants. L'analyse des épreuves sommatives collectées a révélé que ces dispositifs présentent des problèmes de validité étant donné que les outils de mesure (questions et situations) et les outils de jugement (barèmes) investis ne correspondent pas aux outils préconisés par la nouvelle approche (APC). Cela signifie que les pratiques évaluatives des enseignants algériens du FLE ont stagné ; un résultat auquel nous sommes parvenue suite à la comparaison des BEM pré-APC aux BEM post-APC ; une comparaison qui a mis en valeur les similitudes des deux catégories des épreuves où les compétences de compréhension et de production sont évaluées séparément, les ressources linguistiques occupent plus que le taux nécessaire, les situations ne respectent pas les exigences... Ces résultats sont à l'origine de nos propositions qui présentent des situations d'évaluation et une grille de correction, conçues en tenant compte des instructions officielles.

## Mots-clés

Validité, évaluation sommative, dispositifs, FLE.

## ملخص

يركز بحثنا على صلاحية التقييم التجميعي للفرنسية كلغة أجنبية في الجزائر، ويحاول دراسة قدرة أجهزة التقييم التي يعتمد عليها حاليًا مدرسو الطور المتوسط للكشف عن الكفاءات الحقيقية لمتعلميهم. أظهر تحليل الاختبارات النهائية التي تم جمعها أن هذه الأجهزة تحتوي مشاكل في الصلاحية نظرًا لأن أدوات القياس (الأسئلة والوضعيّات) وأدوات الحكم (المقاييس) المستخدمة لا تتوافق مع الأدوات التي أوصى بها النهج الجديد (المقاربة بالكفاءات). هذا يعني أن ممارسات تقييم المعلمين الجزائريين للفرنسية كلغة أجنبية قد ركدت؛ نتيجة توصلنا إليها بعد مقارنة شهادة التعليم المتوسط قبل المقاربة بالكفاءات مع شهادة التعليم المتوسط لما بعد المقاربة بالكفاءات؛ مقارنة أبرزت أوجه التشابه بين فئتي الاختبارات حيث يتم تقييم مهارات الفهم والإنتاج بشكل منفصل، وتحتمل الموارد اللغوية أكثر من المعدل اللازم، والوضعيّات لا تفي بالمتطلبات ... هذه النتائج هي أصل مقترحاتنا التي تعرض حالات التقييم وشبكة تصحيح، مصممة مع مراعاة التعليمات الرسمية.

## الكلمات المفتاحية

الصلاحية، التقييم التجميعي، الأجهزة، الفرنسية كلغة أجنبية

## Abstract

Our research, focused on the validity of the summative evaluation of French as a Foreign Language in Algeria, attempts to study the capacity of the assessment devices currently introduced by middle school teachers to reveal the real skills of their learners. The analysis of the summative tests collected revealed that these devices present problems of validity given that the measurement tools (questions and situations) and the judgment tools (scales) used do not correspond to the tools recommended by the new approach (competence-based approach). This means that the evaluation practices of Algerian teachers of French as a Foreign Language have stagnated; a result we arrived at following the comparison of tests introduced before and after the introduction of this approach; a comparison which revealed the similarities of the two categories of tests where the comprehension and production skills are assessed separately, the linguistic resources occupy more than the necessary rate, the situations do not respect the requirements ... These results are in the origin of our proposals which present evaluation situations and a correction grid, designed taking into account the official instructions.

## Key words

Validity, summative evaluation, devices, French as a Foreign Language.