

Les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants

Analyse du regard des enseignants du primaire

Gonzague Yerly

THÈSE DE DOCTORAT

présentée à la Faculté des lettres de l'Université de Fribourg (Suisse)

**Approuvé par la Faculté des lettres sur proposition des professeurs Dr Jean-Luc
Gurtner (Université de Fribourg) et Dr Dany Laveault (Université d'Ottawa)**

Fribourg, le 29 septembre 2014

Le Doyen, Prof. Dr Marc-Henry Soulet

RÉSUMÉ

Depuis plusieurs décennies, l'évaluation externe des acquis des élèves (EE) connaît une expansion importante dans la plupart des systèmes éducatifs. A sa fonction pédagogique originelle de certification des apprentissages des élèves est venue s'ajouter celle d'un outil politique de pilotage du système. L'EE devrait, théoriquement, permettre d'améliorer la qualité de l'Ecole. Néanmoins, le développement de cet outil et son bienfondé suscitent de nombreuses controverses. Si les effets théoriques de l'EE sont déterminés (bénéfiques ou néfastes), les effets réels, sur le terrain, sont encore peu connus. De plus, si l'impact de ces politiques de *testing* a été souvent évalué au regard des performances des élèves, les recherches sont encore rares au sujet de leur impact sur les acteurs de terrain, notamment sur les enseignants et leurs pratiques. C'est l'objectif de la thèse.

La présente contribution interroge les perceptions des enseignants quant à l'EE, ses effets sur leurs pratiques et les facteurs qui les influencent. Elle fait tout d'abord la recension des travaux théoriques et des recherches empiriques sur le sujet. Une recherche empirique est ensuite menée en deux phases complémentaires (qualitative et quantitative). La recherche s'inscrit dans le contexte de la scolarité primaire du canton de Fribourg (Suisse), qui comprend deux types distincts d'EE (épreuves cantonales de référence et d'orientation). Les analyses portent sur les expériences de huit enseignants interrogés dans des entretiens et sur les réponses de quelques trois cents enseignants à un questionnaire.

Les résultats démontrent premièrement que l'EE est généralement considérée par les enseignants interrogés comme un "mal nécessaire". D'une part, ils peuvent la considérer comme utile pour leurs pratiques, surtout afin de se rassurer. D'autre part, les enseignants interrogés peuvent ressentir une certaine pression face aux épreuves, ayant le sentiment que ce sont davantage leurs propres performances qui sont évaluées, plutôt que celles de leurs élèves. De manière similaire, les enseignants interrogés perçoivent généralement des effets néfastes et des effets bénéfiques des EE sur leurs pratiques. Elle peut les pousser à développer leurs pratiques (développement du professionnalisme des enseignants, encouragement à la collaboration, harmonisation des pratiques d'évaluation sommative), mais peut aussi les entraîner vers des pratiques qui sont en contradiction avec leur modèle pédagogique habituel et/ou idéal (*teaching to the test*, rétrécissement du curriculum enseigné). Un intérêt particulier de notre recherche est d'expliquer ces effets. Les résultats montrent qu'ils ne dépendent pas du profil des enseignants, mais du type d'EE (de référence vs. d'orientation) et de la perception du dispositif par les enseignants. En conclusion, notre travail identifie les conditions qui permettent à l'EE d'être un moyen de perfectionnement continu pour les enseignants.

Mots-clés : évaluation externe des acquis des élèves – scolarité primaire – enseignants – perceptions – pratiques – *mixed method research*

REMERCIEMENTS

La réalisation de cette thèse de doctorat a été rendue possible grâce au soutien de nombreuses personnes à qui j'exprime toute ma reconnaissance.

Au Professeur Jean-Luc Gurtner qui m'a donné l'opportunité de réaliser ce projet et qui m'a soutenu grâce à son regard et ses conseils avisés.

Au Professeur Dany Laveault pour son soutien et ses feedback enrichissants.

Aux collaborateurs du Département des Sciences de l'éducation de l'Université de Fribourg qui m'ont offert leur aide et une ambiance de travail très agréable.

Aux enseignants qui ont répondu à mes questions et ont témoigné de leur précieuse expérience.

Au Service de l'Enseignement Obligatoire de langue Française du canton de Fribourg (SEnOF) qui m'a donné accès au terrain.

Aux étudiants de Master en Sciences de l'éducation de l'Université de Fribourg qui ont contribué à certaines phases du travail.

Aux chercheurs d'horizons divers qui m'ont offert des échanges enrichissants au sujet de mon travail.

A ma famille et à mes proches qui m'offrent l'équilibre d'une vie sereine et qui m'ont aidé et soutenu à chaque instant de ce travail.

TABLES DES MATIÈRES GÉNÉRALE

1. Introduction	1
Chapitre 1 - Cadre théorique	5
2. Définition de l'objet de recherche : l'évaluation externe des acquis des élèves.....	7
3. Fondements et développement de l'usage de l'évaluation externe.....	10
4. Développement et usage de l'évaluation externe dans différents contextes.....	27
5. Fonctions de l'évaluation externe des acquis des élèves	33
6. Effets de l'évaluation externe des acquis des élèves	49
7. Facteurs qui influencent les effets de l'évaluation externe sur les pratiques des enseignants..	75
8. Perceptions de l'évaluation externe par les enseignants.....	83
9. Recommandations pour la mise en place de l'évaluation externe	85
10. Questions de recherche et hypothèses	88
11. Articulation de notre recherche et choix méthodologiques	91
12. Implications théoriques et pratiques de la thèse	92
Chapitre 2 - Contexte de la recherche	93
13. Développement de l'évaluation externe en Suisse	95
14. Développement de l'évaluation externe dans le canton de Fribourg.....	103
Chapitre 3 - Première partie empirique.....	111
15. Objectifs de la première partie empirique	113
16. Méthodologie de la première partie empirique.....	114
17. Présentation des résultats de la première partie empirique.....	119
Chapitre 4 - Seconde partie empirique.....	151
18. Objectifs de la seconde partie empirique.....	153
19. Méthodologie de la seconde partie empirique	154
20. Présentation des résultats de la seconde partie empirique	174
Chapitre 5 - Interprétation des résultats.....	203
21. Perceptions de l'évaluation externe par les enseignants (QR1).....	204
22. Effets de l'évaluation externe sur les pratiques des enseignants (QR2)	209

23. Facteurs qui influencent les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants (QR3)	213
Chapitre 6 - Conclusions.....	219
24. L'EE, un "mal nécessaire" pour les enseignants	220
25. Des effets bénéfiques et néfastes de l'EE.....	222
26. L'usage des résultats de l'EE comme prédicteur principal de ses effets	227
27. L'EE, un moyen de perfectionnement continu pour les enseignants.....	228
28. Limites et prolongements du travail	234
Références	235
Annexes	247

1. Introduction

L'«évaluation externe» est une approche spécifique de l'évaluation qui, à l'inverse de l'évaluation interne, est mandatée et conçue par des personnes extérieures à l'entité évaluée. Dans le champ des Sciences de l'éducation, cet instrument vise une grande variété d'objets tels que les politiques éducatives, les programmes, les établissements, les enseignants¹, les acquis des élèves², etc. La présente thèse, quant à elle, traite exclusivement de l'évaluation externe des acquis des élèves. Cette dernière occupe aujourd'hui une place prédominante dans la plupart des systèmes scolaires mais suscite de nombreux débats, aussi bien théoriques que pratiques. En effet, l'évaluation externe des acquis des élèves est la cible de controverses quant au bienfondé de son usage et quant à ses effets réels sur le terrain.

1.1. Le développement de l'usage de l'évaluation externe des acquis des élèves dans les systèmes éducatifs

L'usage de l'évaluation externe des acquis des élèves n'est pas nouveau dans les systèmes éducatifs. Dans de nombreux contextes européens ou nord-américains, elle tient depuis longtemps le rôle d'uniformiser la certification des élèves à la fin de certains degrés de la scolarité dans un but de promotion ou de sélection. L'évaluation externe des acquis des élèves a connu un développement important dans les systèmes éducatifs des pays de l'OCDE et son expansion s'est accélérée. A sa fonction pédagogique originelle, est venu s'ajouter celle d'outil politique. Elle ne sert aujourd'hui plus seulement à évaluer l'apprentissage des élèves, mais doit permettre d'améliorer la qualité du système dans sa globalité (observation du système et intervention sur le terrain). (Eurydice, 2009; Mons, 2009). Ce développement s'explique par la mutation des systèmes éducatifs et de leur mode de régulation. Lié surtout à des facteurs économiques, la figure des autorités dans les politiques éducatives est passée d'un « Etat éducateur » à un « Etat évaluateur » (Demailly, 2001; Maroy, 2013). L'évaluation externe des acquis des élèves est un des agents principaux de cette nouvelle organisation politique (Broadfoot, 2000; Pons, 2011). L'« Etat évaluateur » est caractérisé par quatre traits principaux : 1) l'école est perçue comme un système de production dans lequel les résultats des élèves sont le principal indicateur de qualité; 2) les contenus curriculaires sont uniformisés et quantifiés (standards de formation) ; 3) l'évaluation externe des acquis des élèves est un moyen privilégié pour mesurer la qualité de l'école (*testing*); 4) les résultats des élèves sont un moyen d'action publique (*accountability*) (Maroy, 2013). Enfin, la globalisation de ce mode de régulation permet des comparaisons internationales, par

¹ Afin de faciliter la lecture du texte, le genre masculin est utilisé comme générique.

² Le terme «acquis» désigne les apprentissages réalisés dans le cadre de la scolarité.

des enquêtes à grande échelle portant sur la performance des élèves (par ex. PISA, PIRLS, TIMMS, etc.). Ces enquêtes sont volontiers prises comme modèle pour la régulation des systèmes, basée aujourd'hui principalement sur l'évaluation externe des acquis des élèves (Mons, 2007a).

1.2. L'évaluation externe des acquis des élèves cible de controverses

Alors que l'évaluation externe des acquis des élèves prend aujourd'hui une place toujours plus importante dans les systèmes éducatifs, elle suscite de nombreuses polémiques sur les plans scientifiques, pratiques et médiatiques. Son apport en tant qu'instrument susceptible de contribuer à l'amélioration la qualité des systèmes éducatifs n'est pas encore prouvé et reste débattu. Les recherches empiriques présentent des résultats bien souvent contradictoires (Mons, 2009).

D'un côté, il est généralement accepté que l'évaluation externe des acquis des élèves peut avoir des effets bénéfiques pour chaque acteur du système. Elle est une source d'informations importantes et utiles pour les élèves (et leurs parents), les enseignants et les autorités. Premièrement, elle assure à tous de bénéficier d'une évaluation "équitable et objective" durant le cursus scolaire (surtout dans un but de promotion et/ou d'orientation). Deuxièmement, elle permet de situer les performances individuelles et collectives par rapport à une norme (interprétation normative) et/ou par rapport à des objectifs des programmes (interprétation critériée). Troisièmement, elle permet d'agir sur les zones critiques (régulation) (Crahay, 2009; Monseur & Demeuse, 2005).

D'un autre côté, cet instrument est la cible de nombreuses critiques. On lui reproche surtout son aspect uniquement quantitatif. Outre des difficultés méthodologiques, c'est surtout le fait de prendre les résultats des élèves comme unique indicateur de qualité qui est remis en cause. Cet indicateur ne permettrait pas d'évaluer l'ensemble d'un système scolaire (Bélaïr, 2005; Duru-Bellat & Jarousse, 2001). L'évaluation externe serait alors la cause d'effets dangereux et néfastes pour tous les acteurs et le système dans sa globalité (Mons, 2009).

1.3. Les effets de l'évaluation externe sur le terrain, besoin de résultats empiriques

Il est attendu que l'évaluation externe des acquis des élèves améliore la qualité du système, notamment par une meilleure performance des élèves mais aussi des enseignants. Si ces effets théoriques de l'évaluation externe sont déterminés, les effets réels, sur le terrain, sont encore peu connus. Des recherches empiriques sont alors nécessaires pour comprendre par quels processus l'évaluation externe peut permettre d'améliorer la qualité des systèmes (Laveault, 2009a ; Maroy, 2012a). De plus, si l'impact de ces politiques de *testing* a été souvent évaluées par des chercheurs au regard des

performances des élèves, les recherches sont encore rares au sujet de leur impact sur les acteurs de terrain, notamment sur les enseignants et leurs pratiques (Pons, 2011).

Les résultats des élèves ont été souvent analysés pour évaluer la pertinence des politiques d'évaluation externe. Ils en sont en effet l'indicateur privilégié, car ils se trouvent en bout de chaîne (*output* final). Toutefois, les recherches montrent des résultats contradictoires. Les chercheurs ne s'accordent pas sur le fait que l'évaluation externe des acquis des élèves soit une plus-value ou non pour les élèves, ni en terme de performance, ni en terme d'égalité des chances (Mons, 2007a; Wößmann, 2007).

En outre, les recherches sur l'impact de ces dispositifs sur les enseignants et leurs pratiques sont encore rares. Pourtant, de leurs pratiques dépend la qualité du système scolaire (par les performances des élèves). Outre l'impact global et politique de l'évaluation externe des acquis des élèves, il est également nécessaire de s'intéresser à son impact au niveau de la classe, de l'usage effectif de ces dispositifs par les enseignants et de ses effets sur leurs pratiques (Lafontaine, Soussi, & Nidegger, 2009; Laveault, 2009b; Mottier Lopez, 2009).

Des recherches empiriques sont alors nécessaires pour répondre à ces questions :

- *Comment l'évaluation des acquis des élèves est-elle perçue par les enseignants ?*
- *Quels usages sont faits de l'évaluation des acquis des élèves par les enseignants ?*
- *Quels sont les effets réels de l'évaluation externe des acquis des élèves sur les pratiques des enseignants ?*
- *Par quels processus fonctionnent les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants ?*
- *Comment promouvoir les effets bénéfiques de l'évaluation externe des acquis des élèves sur les pratiques des enseignants ?*
- *Comment éviter ou réduire les effets néfastes de l'évaluation externe des acquis des élèves sur le terrain ?*

1.4. Articulation du travail

La présente thèse traite des questions soulevées dans les paragraphes précédents. Elle est structurée en six chapitres. Le chapitre 1 présente les concepts théoriques nécessaires à la bonne compréhension du sujet d'étude et fait la recension des études empiriques menées sur le sujet dans divers contextes. Ces apports permettent d'amener de premières réponses aux questions posées dans l'introduction, puis de formuler les questions qui nécessitent davantage de résultats empiriques. Notre recherche y répond en interrogeant des enseignants grâce à deux méthodes complémentaires (qualitative et quantitative). Le

chapitre 2 présente le contexte de l'étude, la scolarité primaire du canton de Fribourg (CH), qui comprend deux types d'évaluation externe des acquis des élèves. Les chapitres 3 et 4 présentent séparément les deux phases de notre enquête sur le terrain. Pour chacune, nous présentons la méthode utilisée et les résultats. Le chapitre 5 fait la synthèse des résultats de ces deux phases afin de répondre à nos questions de recherche. Dans le chapitre 6, les résultats sont discutés, en les confrontant notamment à la littérature. En conclusion, notre travail identifie les conditions qui permettent à l'EE d'être un moyen de perfectionnement continu pour les enseignants.

CHAPITRE 1

CADRE THÉORIQUE

2. Définition de l'objet de recherche : l'évaluation externe des acquis des élèves	7
2.1. L'évaluation externe et ses différents objets	7
2.2. L'évaluation externe des acquis des élèves : définitions	7
2.3. L'évaluation externe des acquis des élèves : notre définition	9
3. Fondements et développement de l'usage de l'évaluation externe	10
3.1. L'évaluation externe, outil d'évaluation uniformisée des apprentissages des élèves	10
3.2. L'évaluation externe, nouvel outil de pilotage des systèmes éducatifs	10
3.3. L'évaluation externe, outil ancré dans les mutations des systèmes éducatifs.....	11
3.4. L'évaluation externe, outil de l'Etat évaluateur	15
3.5. Les critiques du modèle de l'Etat évaluateur.....	24
4. Développement et usage de l'évaluation externe dans différents contextes	27
4.1. L'évaluation externe en Angleterre	27
4.2. L'évaluation externe aux Etats-Unis	28
4.3. L'évaluation externe en France	29
4.4. L'évaluation externe en Allemagne.....	30
4.5. L'évaluation externe au Canada	30
4.6. L'évaluation externe en Belgique francophone.....	31
5. Fonctions de l'évaluation externe des acquis des élèves	33
5.1. Les fonctions multiples de l'évaluation externe	33
5.2. Les fonctions politiques de l'évaluation externe	35
5.3. Les fonctions pédagogiques de l'évaluation externe	40
5.4. Les tensions entre les différentes fonctions de l'évaluation externe	45

6. Effets de l'évaluation externe des acquis des élèves.....	49
6.1. Les effets de l'évaluation externe sur le système éducatif.....	49
6.2. Les effets de l'évaluation externe sur les enseignants	54
6.3. Les effets de l'évaluation externe sur les cadres intermédiaires.....	72
6.4. Les effets de l'évaluation externe sur les élèves.....	73
6.5. Les effets de l'évaluation externe sur les parents	74
7.Facteurs qui influencent les effets de l'évaluation externe sur les pratiques des enseignants ...	75
7.1. Les facteurs liés au contexte	79
7.2. Les facteurs liés à l'épreuve externe.....	79
7.3. Les facteurs liés aux enseignants.....	80
8. Perceptions de l'évaluation externe par les enseignants	83
8.1. Les enseignants plutôt favorables.....	83
8.2. Les résistances des enseignants	84
9. Recommandations pour la mise en place de l'évaluation externe.....	85
10. Questions de recherche et hypothèses.....	88
11. Articulation de notre recherche	91
12. Implications théoriques et pratiques de la thèse	92

2. Définition de l'objet de recherche : l'évaluation externe des acquis des élèves

L'objectif de ce premier point est de mieux définir notre objet de recherche, "l'évaluation externe des acquis des élèves". Nous présentons tout d'abord comment elle est définie dans la littérature, puis nous en formulons notre propre définition.

2.1. L'évaluation externe et ses différents objets

Dans le champ des Sciences de l'éducation, le terme "évaluation externe" détermine une approche spécifique de l'évaluation qui, à l'inverse de l'évaluation interne, est mandatée et conçue par des personnes extérieures à l'entité évaluée. Ces deux approches se différencient également par leurs démarches :

(L'évaluation externe) se focalise sur l'analyse objective d'éléments factuels en adoptant un regard extérieur aux situations observées. [...] L'approche interne privilégie le vécu, le sens et la subjectivité de l'objet, alors qu'avec l'approche externe, ce sont la rigueur, la mesure, et l'objectivité qui gouvernent la démarche. L'évaluation externe est également caractérisée par une dimension quantitative mobilisant l'analyse de données empiriques, ce qui a pour conséquence la simplification de la réalité observée (Suchaut, 2008, p. 89).

Même si elles sont généralement caractérisées par une dimension quantitative, les démarches de l'évaluation externe sont multiples et parfois complémentaires (quantitatives et qualitatives). Elles dépendent surtout des objets évalués. Ceux-ci sont d'ailleurs nombreux et peuvent être interdépendants : les systèmes éducatifs, les politiques éducatives, les programmes, les établissements, les enseignants, les acquis des élèves... La présente thèse traite spécifiquement de l'évaluation externe des acquis des élèves³.

2.2. L'évaluation externe des acquis des élèves : définitions

Même si ses objets peuvent être divers, les recherches consacrées à "l'évaluation externe" portent pour la plupart sur des dispositifs destinés à l'évaluation des acquis des élèves. Si cette spécification est bien souvent sous-entendue, d'autres formes d'évaluation externe (sous forme d'audits ou d'inspections) ne portent pas sur les apprentissages des élèves mais plutôt sur des modes d'organisation ou des processus éducatifs (gestion des ressources, tâches éducatives, relations ...) (Maroy, 2012).

³ Le terme "acquis" désigne les apprentissages réalisés dans le cadre de la scolarité.

De nombreux auteurs donnent une définition de l'évaluation externe des acquis des élèves. En voici quelques-unes, issues de travaux élaborés dans différents pays :

- « A l'opposé des évaluations assurées par les enseignants et dont la portée est interne à la classe, les épreuves externes sont élaborées par l'administration scolaire et concernent le plus souvent tous les élèves du pays ou du canton » (Wirthner & Ntamakiliro, 2008, p.91).
- « External testing are those controlled and/or mandated by an external authority » (Haney & Madaus, 1986, p.9)
- « Evaluation réalisée par des personnes ne faisant pas partie de l'équipe éducative chargée de réaliser le programme (*external evaluation* en anglais)» (De Landsheere, 1979, p.113).
- « Elle (l'évaluation externe) est toujours construite et parfois administrée par une personne extérieure ou un organisme externe à la classe » (Dierendonck, 2008, p.5).

Malgré les différents contextes dont sont issus ces définitions, elles partagent des points communs : il s'agit d'une mesure à large échelle (pays, région, degré scolaire, classe d'âge...) des acquis (synonyme d'apprentissages) des élèves. Elle est mandatée et élaborée par une entité externe à la classe. Toutefois, dans la littérature, certains points de divergence apparaissent quant au rôle de l'enseignant de la classe dans l'administration (passation et correction) de l'évaluation externe. Pour certains auteurs, l'évaluation a un caractère « mixte » (interne et externe) si l'enseignant administre et corrige lui-même les épreuves (Monseur & Demeuse, 2005 ; Haney & Madaus, 1986). C'est par exemple le cas, des nombreux tests commercialisés aux Etats-Unis, ou de ce qui est appelé "évaluation externe" en Communauté française de Belgique. Toutefois, comme elles portent sur un échantillon large, les conditions d'administration et de correction doivent être uniformisées. On parle alors d'évaluations externes standardisées (Wirthner & Ntamakiliro, 2008). De Landsheere (1979), dans son dictionnaire de l'évaluation, donne une définition de la standardisation d'un test :

Définition précise des modalités d'un test. [...] Tous les élèves d'une même catégorie sont soumis à une même épreuve. En outre, les conditions d'administration et de correction sont uniformisées. Idéalement, pour que les conditions soient comparables, le test devrait toujours être subi dans des conditions identiques (p.248).

Cet auteur donne également les conditions à respecter pour une standardisation d'une épreuve, présentées dans le tableau 1. De Landsheere (1979) ajoute cependant que ces conditions sont idéales et qu'une telle uniformité est difficilement réalisable. Mais, il ajoute qu'il est important de s'en approcher.

TABLEAU 1 – Conditions pour la standardisation d'un test, De Landsheere (1979, p.248)

- | |
|--|
| <ul style="list-style-type: none">a) Même jour, heure, degré initial de fatigue, température, dispositions matérielles (sièges, etc.), silence, etc. ;b) Même degré de familiarité des sujets avec les tests, en général, et la technique utilisée, en particulier : un élève habitué depuis longtemps aux réponses à choix multiples ou à l'utilisation de feuilles-réponses standardisées est avantagé par rapport à un autre enfant qui ignore tout de ces procédés ;c) Même motivation ;d) Mêmes consignes initiales et même entraînement ;e) Même durée, si celle-ci est limitée. |
|--|

2.3. L'évaluation externe des acquis des élèves : notre définition

Les différents éléments exposés nous permettent de clarifier notre objet de recherche et d'en formuler notre propre définition.

L'évaluation externe des acquis des élèves est un dispositif d'évaluation des apprentissages (ou acquis) des élèves qui est mandaté et conçu par des personnes extérieures à la classe (externe) et qui est destiné à tous les élèves (ou à un échantillon représentatif) d'un même système scolaire (pays, région, etc.) et d'un même degré pour qui les conditions d'administration (contenu, passation et correction) sont uniformes (standardisée).

Dans la thèse, pour éviter les répétitions, certains synonymes du terme évaluation externe standardisée des acquis des élèves sont utilisés : évaluation externe, évaluation standardisée, épreuve externe, examen externe, test externe. Le sens des adjectifs externe et standardisé est maintenu, même s'il n'apparaît pas. De plus, pour faciliter la lecture du texte, l'abréviation EE est utilisée.

3. Fondements et développement de l'usage de l'évaluation externe

Le présent point expose les fondements de l'usage de l'évaluation externe des apprentissages des élèves dans les systèmes éducatifs et les raisons de son expansion au cours des dernières décennies, en lien avec l'évolution des systèmes éducatifs.

3.1. L'évaluation externe, outil d'évaluation uniformisée des apprentissages des élèves

Selon le rapport Eurydice (2009), l'usage de l'évaluation externe des acquis des élèves dans les systèmes éducatifs européens n'est pas nouveau. Historiquement, ces tests avaient pour but d'uniformiser la certification des élèves à la fin de certains degrés de la scolarité (comme les examens du *Primary Certificate* en Irlande dès 1967 ; l'examen de fin de la *Folkskole* au Danemark dès 1975). En outre, les évaluations externes étaient également destinées à assurer des conditions équitables pour le passage des élèves à un degré d'étude supérieur (notamment aux Pays-Bas et au Luxembourg dès 1968). Aux Etats-Unis également, les évaluations externes standardisées servaient à la promotion et à la sélection des élèves dès le XIX^{ème} siècle (Carnoy & Loeb, 2002)⁴.

3.2. L'évaluation externe, nouvel outil de pilotage des systèmes éducatifs

On assiste à un basculement dans l'usage de l'évaluation externe à la fin des années quatre-vingt. A sa fonction pédagogique précédemment exposée (au niveau micro) est venu s'ajouter celle d'un outil de politique éducative (au niveau macro). Elle ne sert alors plus uniquement à la certification et/ou à l'orientation des élèves mais aussi (et surtout) à la gestion et à l'orientation du système scolaire entier.

[...] elles (les épreuves externes) sont non seulement destinées à évaluer les apprentissages des élèves mais aussi, sinon davantage, à estimer le niveau de la classe, de l'école, du canton, du pays (Tessaro & Ntamakiliro, 2010a, p. 1).

Les résultats de l'évaluation externe sont devenus un outil d'information qui sert au « pilotage » et à la « régulation » du système :

- *Pilotage*. Les résultats des élèves issus du niveau microscopique (celui de la classe ou de l'établissement) servent aux prises de décisions au niveau macroscopique (celui des politiques éducatives (Landsheere, 1994).

⁴ Les différentes fonctions pédagogiques et politiques de l'évaluation externe sont exposées dans le point 5 (p.33-48).

- *Régulation.* La régulation fait partie des outils de pilotage. Dans le champ des politiques éducatives⁵, elle est entendue « [...] au sens large comme le processus de production de règles et d'orientation des conduites des acteurs » (Maroy & Dupriez, 2000, p. 73). Les informations fournies par l'évaluation externe des acquis des élèves servent donc à produire des règles, à les contrôler et à orienter les « conduites » des acteurs du système vers celles-ci.

Les nouvelles fonctions politiques de l'évaluation externe trouvent leurs origines dans les mutations des politiques publiques lors des dernières décennies. Des changements sociaux et économiques ont obligé les Etats à revoir les objectifs de leurs politiques publiques et leurs modes de régulation (Maroy, 2005a). Ceux-ci sont exposés dans le point suivant.

3.3. L'évaluation externe, outil ancré dans les mutations des systèmes éducatifs

Se référant à la recherche européenne Reguleduc⁶, Maroy (2005) observe une évolution semblable des systèmes éducatifs européens et nord-américains. Tout en gardant des particularismes nationaux et/ou régionaux, le modèle de régulation « bureaucratique-professionnel » en vigueur dans les années cinquante et soixante s'est transformé et a laissé place à un modèle « post-bureaucratique ». Nous décrivons cette évolution afin de comprendre les usages actuels de l'évaluation externe.

3.3.1. Le modèle de régulation bureaucratique-professionnel : l'Etat éducateur

La plupart des systèmes scolaires ont été construits dès le XIX^{ème} siècle autour de la notion d'un « Etat éducateur ». Celui-ci était responsable d'offrir un enseignement « à la masse ». Cette offre éducative représentait un progrès économique et social (Maroy & Dupriez, 2000). Bien entendu, cette mise en place s'est faite avec des différences nationales ou régionales. Cependant, deux caractéristiques principales étaient partagées dans leur modèle de régulation, justement qualifié de « bureaucratique-professionnel » (Maroy, 2005a):

- *Bureaucratique.* Dans ce modèle, les règles étaient uniformisées dans le but d'assurer l'égalité de traitement des élèves et l'accès de tous à l'éducation. Le système scolaire était alors réglementé et organisé de manière hiérarchique.

⁵ Le terme « régulation », utilisé ici dans le champ des politiques éducatives, se différencie selon les auteurs et leur approche. Nous reprenons cette distinction au point 5.2.2 (p.36-37).

⁶ Cette recherche financée par l'UE a pour objet d'étude l'évolution des modes de régulation et de gouvernance des systèmes éducatifs européens.

- *Professionnel*. Dans ce modèle, l'aspect bureaucratique des systèmes éducatifs était associé à une grande autonomie laissée aux professionnels de l'enseignement (enseignants et inspecteurs), acquise grâce à leur expertise de la profession.

En résumé, le modèle bureaucratico-professionnel était un système « normatif » dans lequel les acteurs avaient une « obligation de moyens » (Demailly, 2001).

C'est le paradigme normatif qui régulaient les pratiques des cadres et des enseignants. Son contenu était la priorité accordée à l'utilisation des bons moyens, des bonnes manières de faire, ou autrement dit la conformité de la pratique professionnelle à des normes définies par des textes officiels, des programmes nationaux et des traditions professionnelles (elles étaient censées produire, par principe, les bons résultats) (p.16).

Dans ce modèle, les normes pédagogiques et leur régulation étaient dirigées de manière conjointe par l'Etat et les enseignants. Outre une logique institutionnelle et politique (*top-down*), la régulation des systèmes éducatifs était menée simultanément par les acteurs du système (*bottom-up*).

La régulation est donc un processus multiple par ses sources, ses mécanismes, ses objets, mais aussi par la pluralité des acteurs qui la construisent (au niveau transnational, national, local). Dans les faits la régulation est toujours une multirégulation complexe, parfois conflictuelle et potentiellement contradictoire (Maroy, 2005b, p. 6).

Il est important de noter que, dans ce modèle, les autres acteurs de l'éducation (parents, élèves) n'avaient pas voix de citer. Pour conclure, ce modèle général est à nuancer par des particularités importantes entre les divers systèmes éducatifs, surtout au niveau de leur centralisation. En effet, certains pays comme la Belgique ou l'Angleterre accordaient des initiatives importantes aux niveaux locaux et/ou religieux (Maroy, 2005b).

3.3.2. Le modèle de régulation post-bureaucratique : d'un Etat éducateur à un Etat évaluateur

Selon Mons (2007 & 2009) et Pons (2011), les systèmes éducatifs connaissent (aujourd'hui encore) de nombreux changements depuis les années 1970 en Amérique du Nord et les années 1980-1990 en Europe. Ces périodes furent marquées par des changements économiques importants tels que la stagflation ou la mondialisation. Ceux-ci ont fait naître un courant de pensée néo-libérale qui remet en question la place traditionnelle de l'Etat dans la société. Il s'agit du *New Public Management (NPM)*. Celui-ci est apparu aux USA, puis s'est développé en Europe. Son objectif prioritaire est de freiner les dépenses publiques, de rationaliser l'action des administrations. Les réformes ancrées dans cette mouvance guident l'Etat vers un meilleur contrôle coût-efficacité de ses actions, en suivant les principes énoncés ici (Mons, 2009 ; Pons, 2011) :

- a) Les services publics doivent être mesurables et comparables (*benchmarking*) ;
- b) Leur mesure doit être faite grâce à des outils scientifiquement valides (*testing*) ;

- c) Les acteurs publics bénéficiant de davantage d'autonomie sont redevables au système (*accountability*);
- d) L'organisation publique doit être régulée par les résultats probants (*evidence-based Policy*).

Cette tendance générale touche également les politiques éducatives, créant des débats publics sur l'école. Selon les principes du *NPM*, l'Etat doit garantir que son système éducatif est efficient. Dans ce sens, sont apparues de nombreuses réformes. On peut tout d'abord citer le rapport américain *A nation at risk* de 1983 ou les *Black Papers* anglais des années 1980. Plus récemment, une deuxième vague de réformes telles que l' *Education Reform Act* de 1988 en Angleterre, le décret *Missions* en Belgique en 1997, les lois sur la décentralisation en France dès 1989 ou la loi fédérale américaine de 2001 *No Child Left Behind* (Eurydice, 2009; Maroy & Dupriez, 2000; Mons, 2009). Nous pouvons également citer la réforme *HarmoS* en Suisse qui est le vecteur de changements importants correspondant aux réformes citées⁷.

Bajomi et Barroso (2002, cités par Maroy, 2005b) voient six tendances vers lesquelles convergent ces changements institutionnels :

- a) *Une plus grande autonomie des établissements*. De nombreuses responsabilités sont laissées aux établissements scolaires. Cette tendance est observée dans la plupart des systèmes scolaires. Cette autonomie laissée aux établissements scolaires vient du mouvement du *School Improvement* qui considère que l'amélioration de l'efficacité et de la qualité de l'école repose en partie sur la mobilisation et l'implication des enseignants au sein des établissements scolaires. Elle suit également la thèse de l' « organisation apprenante » qui entend valoriser le développement des compétences de l'organisation pour la rendre davantage efficace (Normand, 2005).
- b) *La recherche d'un équilibre entre centralisation et décentralisation*. Les Etats historiquement centralisés ont tendance à chercher davantage à décentraliser le pouvoir décisionnel à des instances intermédiaires ou locales (par exemple : France, Portugal, Hongrie) alors que les Etats historiquement fortement décentralisés ont tendance à renforcer leur uniformité, notamment en adoptant des objectifs curriculaires communs (Belgique, Angleterre, Suisse).
- c) *La promotion ou l'assouplissement du choix de l'école par les parents*. On observe dans beaucoup de systèmes scolaires une volonté d'assouplir les règles au sujet du libre choix de l'école par les parents. On parle même dans certains contextes (Communauté francophone de Belgique et Angleterre) de « quasi-marché », les écoles étant en concurrence.

⁷ Cette réforme est présentée dans le point 13.1 (p.95-98).

- d) *La diversification de l'offre scolaire.* Le curriculum étant devenu standard dans les systèmes scolaires, les établissements se démarquent parfois par des offres de services spécifiques (horaires aménagés, classes européennes, spécialisation sur certains domaines).
- e) *L'augmentation du contrôle du travail enseignant.* Les enseignants sont soumis à des formes de contrôle de leur travail de plus en plus importantes (formation continue, présence de cadres de proximité dans les établissements, obligation du travail en équipe).
- f) ***La montée de l'évaluation externe.*** L'évaluation externe devient un outil majeur dans les systèmes scolaires, aussi bien au niveau central que local. Même si ces objets et ces méthodes sont variés, la mesure principalement faite est celle de la performance des élèves.

Ces six aspects, en rupture avec le modèle « bureaucratique-professionnel », conduisent les systèmes scolaires vers un nouveau modèle de régulation « post-bureaucratique » (Maroy, 2005b). Dans ce nouveau modèle, la notion normative d'« obligation de moyen » est supplantée par celle de « d'obligation de résultats » (Demailly, 2001). Toutefois, ils gardent un aspect « bureaucratique » car les règlements et les lois restent la base des politiques, l'Etat étant encore le garant du maintien des règles.

[...] la valorisation des résultats, de la recherche d'efficacité (pouvant aller jusqu'à l'obligation de résultats) est privilégiée par rapport au respect de la règle de droit. La rationalité reste valorisée mais elle est de plus en plus réduite à la rationalité instrumentale (Maroy, 2005b, p. 11).

Maroy (2005b) distingue deux variantes du modèle « post-bureaucratique » qui s'imposent aujourd'hui dans les systèmes éducatifs. Elles sont toutefois combinées dans certains pays (en Angleterre et aux USA notamment):

- « *Etat évaluateur* » (appelé aussi de gouvernance par les résultats ; Policy Evaluation en anglais). Ce modèle fait l'objet d'un développement particulier dans le prochain point (3.4.) car il est l'un des fondements du rôle accru de l'évaluation externe dans les systèmes éducatifs.
- « *Quasi-marché* ». Ce modèle est lié au libre choix de l'école par les parents, ce qui encourage la compétition entre les établissements auxquels il est laissé une grande autonomie pour atteindre les objectifs d'un programme standard centralisé mais qui reçoivent des financements selon leur performance ou leur public. Ce modèle néo-libéral est plutôt anglo-saxon (Angleterre, USA, Australie, Nouvelle-Zélande).

3.4. L'évaluation externe, outil de l'Etat évaluateur

L'expansion de l'usage de l'évaluation externe dans les systèmes éducatifs s'explique par le développement du modèle de l'Etat évaluateur (nommé aussi pilotage par les résultats). Comme exposé auparavant, il apparaît dans une logique de recherche d'efficacité et d'efficience des systèmes éducatifs. L'objectif des réformes est donc d'améliorer la qualité de l'école, tout en rationalisant ses coûts. On fait le pari que l'autonomie des établissements est la clé de cette qualité. Selon Neave (1988, cité par Maroy & Dupriez, 2000), l'Etat s'approprie le contrôle stratégique des systèmes scolaires en définissant des objectifs et les critères de résultats. Cet Etat évaluateur laisse la gestion des moyens d'atteindre ces objectifs à des instances décentralisées et/ou locales (département, municipalités, établissement). Toutefois, l'Etat garde le contrôle du système par des moyens de contrôle (pilotage).

Le "modèle de gouvernance par les résultats" suppose que les objectifs et programmes à réaliser par le système d'enseignement soient définis de façon centrale et que les unités d'enseignements aient une importante autonomie pédagogique et/ou financière. Par ailleurs, ces dernières sont soumises à des contrats. L'Etat central négocie avec les entités locales (comme les établissements scolaires) des "objectifs à atteindre" et lui délègue des responsabilités et des moyens accrus pour réaliser ces objectifs [...]. Par ailleurs, un système d'évaluation externe des performances des établissements et un système d'incitants symboliques ou matériel, voire de sanction, se met en place pour favoriser la réalisation ou l'amélioration des performances du contrat passé par l'Etat et les établissements (ou d'entités d'un niveau supérieur) (Maroy, 2004, p.27).

Pour Maroy (2013), le modèle de l'Etat évaluateur se différencie quant à ses finalités et à ses outils selon le contexte politique et social dans lequel il s'inscrit. Toutefois, l'auteur distingue quatre traits « transversaux » (qui sont d'ailleurs semblables aux quatre principes du *NPM* cités auparavant) :

- 1) L'école est conçue comme un système de production (l'obligation de résultats) ;
- 2) Les résultats du système sont uniformisés et quantifiables (les standards de formation) ;
- 3) Les résultats du système sont mesurables (**l'évaluation externe**) ;
- 4) les résultats du système servent à l'action publique (les politiques d'*accountability*).

Nous développons ces quatre points dans les parties suivantes, car ils sont les fondements de l'usage de l'évaluation externe dans les systèmes éducatifs. A ces points, nous en ajoutons un cinquième :

- 5) les comparaisons internationales (un modèle de régulation par les résultats).

3.4.1. *L'obligation de résultats : l'école conçue comme un système de production*

Dans le modèle de l'Etat évaluateur, l'école est vue comme un « système de production » et non plus comme une « institution » (Maroy, 2005a). A la manière d'une entreprise, les systèmes éducatifs

suivent une logique de résultats. Les politiques menées devant être efficaces, il est nécessaire de comparer *inputs* (investissements) et *outputs* (résultats). C'est l'idée défendue par le courant de la *school effectiveness*, qui s'inspire des théories économiques du « capital humain » et de la « théorie des organisations » (Mons, 2009 ; Maroy, 2013 ; Normand, 2005). On cherche ainsi à expliquer et à améliorer la qualité de l'école par la mesure de ses résultats. C'est à partir des années 1980, que l'on invoque ces modèles économiques dans le but d'améliorer la qualité de l'école (Suchaut, 2008).

L'organisation de l'école a ainsi été réduite à sa dimension managériale et productive, et l'apprentissage à sa dimension mesurable et objective (Normand, 2005, p. 70).

On assiste alors à des mutations dans les politiques scolaires. Contrairement au passé, les résultats souhaités sont formulés en termes de performance individuelle et non en termes de valeurs collectives (Maroy & Mangez, 2011). Les systèmes éducatifs sont ainsi passés de l'« obligation de moyens » (obligation pour les acteurs de se conformer à des pratiques normatives, c'est-à-dire aux « bonnes » pratiques qui étaient censées produire les bons résultats) à l'« obligation de résultats » (Demailly, 2001). Pour l'auteure, il est toutefois dangereux de séparer résultats et moyens. Les programmes doivent encore aiguiller les enseignants dans leur mission éducative, afin d'éviter des dérives. En effet, tous les moyens ne sont pas acceptables pour arriver au résultat attendu (tricherie, exclusion d'élèves...). Maroy et Dupriez (2000) observent l'apparition, dans les écoles, de nouveaux moyens qui traduisent cette nouvelle philosophie :

De manière progressive, depuis quelques dizaines d'années, l'enseignement est sorti d'une logique de vocation où les qualités du maître n'étaient par définition ni mesurables, ni évaluables. L'univers scolaire est en train d'intégrer une série de références nouvelles telles que socle de compétences, profils de formation, épreuves d'évaluation externes, évaluation de compétences, savoir ajouté, coût moyen par élève,.... Et la nouveauté n'est pas que dans le vocabulaire. Plus fondamentalement, il s'agit de faire rentrer les tâches d'enseignement dans un modèle d'ingénierie sociale où les objectifs à atteindre sont définis à l'aide d'indicateurs les plus précis possibles et susceptibles de faire l'objet d'une évaluation, voire d'une mesure externe (p.75).

L'école, pour prouver son bon fonctionnement, est donc contrainte de produire des résultats. Et ce sont les résultats des élèves qui sont l'indicateur privilégié pour déterminer la qualité des systèmes scolaires, étant des outputs mesurables situés au bout de la « chaîne de production ». Pour prouver la performance du système entier, les élèves doivent atteindre des objectifs préalablement fixés, l'évaluation externe en constituant le moyen de mesure.

La variabilité des résultats des élèves apparaît comme la notion centrale et deux concepts clés s'y rattachent : l'efficacité et l'équité. L'efficacité mesure la relation entre les objectifs du système et les résultats observables sur les élèves et leur parcours scolaires. Ces résultats peuvent s'apprécier selon différentes dimensions complémentaires : acquisitions, réussite aux examens, insertion professionnelle... L'équité peut être considérée comme la réduction des disparités entre les résultats des élèves. (Suchaut, 2008, p. 90)

3.4.2. *Les standards de formation : moyen d'harmonisation des contenus curriculaires et indicateurs de qualités des systèmes éducatifs*

La standardisation des contenus curriculaires est un élément majeur dans le modèle de l'Etat évaluateur. Les contenus doivent être uniformes (centralisation) mais une grande autonomie est laissée aux acteurs de terrain (pouvoirs locaux, établissements, enseignants) dans la manière de les atteindre (décentralisation) (Eurydice, 2009 ; Maroy, 2004 ; Mons & Pons, 2006). Il est donc nécessaire que les contenus et/ou les objectifs à atteindre soient les mêmes pour tous à des fins de mesure, notamment pour permettre des comparaisons et une concurrence. Il s'agit là du concept économique de *benchmarking* (Pons, 2011).

Aujourd'hui, dans la plupart des systèmes scolaires de l'OCDE, les contenus curriculaires ont été uniformisés et sont formulés sous forme de "standards de formation". Selon Normand (2005), le terme *standard* (qui vient de l'anglais) n'a pas d'équivalent en français. Il définit le niveau de performance que doivent atteindre les élèves à certains moments de leur cursus. Le terme de standard est apparu pour la première fois en Angleterre vers 1860. Il fixait déjà un niveau d'atteinte attendu pour les élèves dans différentes branches à des fins de classement des élèves selon leurs résultats. Ces standards furent toutefois abolis à cause des effets pervers qu'il eurent notamment sur les pratiques des enseignants (Mons & Pons, 2006). Normand (2005) et Meuret (2012) citent notamment les critiques de Dewey et Haley aux USA concernant la rigidité de ce système qui ne favorise pas les expériences utiles à l'apprentissage. Pour Behrens (2005), les standards étant issus de la pédagogie par objectifs (notamment développée par Bloom), ils sont pour certains une régression et entraînent des réticences influencées par le modèle socioconstructiviste.

Toutefois, les standards sont réapparus dans les systèmes éducatifs pour leur valeur d'indicateurs mesurables et objectifs, valeur prônée par le *NPM* (Normand, 2005). Ils ont pour but de donner une vision claire des objectifs que les élèves doivent atteindre.

L'élaboration de standards de formation représente donc des enjeux importants. Ce processus se place à la fois dans une perspective d'efficacité et dans une logique d'harmonisation. Il innove en rendant les enseignants responsables des acquisitions des élèves [...]. Enfin, il introduit une dynamique de réforme par une logique d'assurance-qualité et suggère la maîtrise administrative de la mécanique scolaire, au risque de restreindre l'autonomie professionnelle des enseignants (Behrens, 2005, p. 6).

Mais surtout, l'atteinte des standards de formation doit être évaluée. Pour Lam (2004, cité par Behrens, 2010) les réformes soutenues par l'élaboration de standards suivent quatre étapes principales :

- a) Définition de standards de performance pour les élèves ;
- b) Responsabilisation des écoles et des enseignants quant à l'atteinte de ces standards ;

- c) **Évaluation externe** à large échelle pour accélérer les changements dans les écoles, soutenue par des politiques d'encouragement (*accountability*) ;
- d) Utilisation des résultats de cette évaluation pour juger de la pertinence des réformes engagées et les accompagner.

Nous montrons dans le point suivant que l'évaluation externe des acquis des élèves s'est développée comme l'outil le plus généralement utilisé, voire l'unique, pour évaluer l'atteinte des standards de formation.

3.4.3. *L'évaluation externe : moyen privilégié pour mesurer la qualité des systèmes éducatifs*

Dans le modèle de l'Etat évaluateur, les acteurs de terrain bénéficiant d'une grande autonomie de moyens, il est nécessaire de contrôler la qualité et l'efficacité du système et d'en déterminer les priorités (pilotage et régulation du système). Pour ce faire, les autorités doivent trouver des moyens qui permettent de mesurer les objectifs à atteindre (standards de formation). Pour de nombreux observateurs, l'évaluation (*testing ou Evaluation Policy*) est devenue l'outil central des politiques éducatives, le moyen privilégié qui va permettre de mesurer et d'analyser la qualité du système (Broadfoot, 2000 ; Demailly, 2001 ; Maroy, 2005 ; Mons & Pons, 2006).

[...] il existe un élément commun à toutes ces réponses spécifiques, un mécanisme qui assure aux gouvernements le pouvoir de pilotage dont ils ont besoin pour parvenir à ce que des milliers d'établissements, plus ou moins indépendants, restent sous leur tutelle. Ce mécanisme, c'est selon moi, l'évaluation (Broadfoot, 2000, p.43).

Se propage alors dans les systèmes éducatifs une « culture de l'évaluation », c'est-à-dire le développement d'outils et de pratiques, mais aussi d'une idéologie générale qui donne à l'évaluation une place capitale (Thélot, 2002). Cet attrait des autorités pour l'évaluation s'explique par des raisons qui vont bien au-delà du contexte scolaire.

Depuis trois décennies environ, l'ensemble des secteurs de l'action publique dans les pays de l'OCDE sont submergés par le déferlement d'une rhétorique considérant l'évaluation comme l'outil cardinal de la bonne gouvernance. Cette idéologie ou « culture de l'évaluation » se déploie autour de l'idée noyau d'une nécessaire intégration des évaluations scientifiques dans la formulation des politiques publiques. (Mons & Crahay, 2011, p.77)

Les décisions de politiques éducatives doivent être prises sur la base de preuves, produites par l'usage d'évaluations scientifiques (*evidence-based Policy*). Pour être scientifique et opérationnelle, l'évaluation des politiques éducatives doit respecter certaines règles. Pour Thélot (2002), pour être de qualité, l'évaluation des politiques éducatives doit être externe. Ceci signifie que l'évaluateur doit être distant de l'évalué. Il ajoute aussi que l'évaluateur doit être « indépendant, puissant, légitime et avoir

accès aux sources ». Dans cette optique, des organismes nationaux ou régionaux d'évaluation ont été créés dans certains pays ou région (France, Royaume-Uni, USA, Suède, Ontario).

L'évaluation doit donc être externe, mais celle-ci peut prendre différentes formes : qualitatives (inspections, audits, *monitoring*) ou quantitatives (résultats des élèves) (Buisson-Fenet & Pons, 2011). Parmi ces diverses formes, celle qui tend à se développer le plus dans le secteur scolaire est quantitative, il s'agit de l'évaluation externe standardisée des acquis des élèves. Comme exposé auparavant, les résultats des élèves, "outputs" observables qui se situent à la fin de la chaîne de production, sont considérés comme le meilleur indicateur, voire la « mesure essentielle », pour évaluer la qualité du système dans sa globalité (Broadfoot, 2000 ; Maroy, 2005; Normand, 2005). Si cette forme d'évaluation externe s'est le plus développée c'est aussi car elle est « en lien direct au moins formellement avec des fonctions traditionnelles de l'école » (Mons & Crahay, 2011, p. 84). En effet, comme exposé auparavant, l'évaluation externe des acquis des élèves tient un rôle pédagogique déjà ancien dans les systèmes éducatifs. Les fonctions que lui octroient les autorités se sont donc multipliées⁸.

Les évaluations externes des acquis des élèves ne sont pas neuves dans les systèmes éducatifs mais elles remplissaient surtout des objectifs de nature pédagogique. [...] Mais elles constituaient rarement pour autant des bases d'informations pour la conduite, plus informée des politiques éducatives (Maroy, 2013, p.20).

L'emploi de l'évaluation externe des acquis des élèves comme outil politique vient des pays anglo-saxons, plus particulièrement des USA. Ceux-ci ont imposé à l'école un modèle managérial prônant la comparaison des résultats et la compétition entre Etats et/ou écoles. L'Etat fédéral a alors imposé la standardisation des contenus et leur évaluation standardisée par un test national (le *Nation's Report Card* appelé autrefois le *National Assessment of Educational Progress (NAEP)*). Les résultats de ce test standardisé donne lieu à des comparaison auxquelles sont liées des récompenses et des sanctions pour les établissements et les enseignants (Linn, 2000 ; Meuret, 2012 ; Normand, 2005). Puis, l'évaluation externe des acquis des élèves est devenue un outil privilégié des politiques éducatives de manière généralisée dans les pays de l'OCDE, devenant une nécessité pour le pilotage des systèmes éducatifs nationaux, locaux ou internationaux (Demailly, 2001 ; Maroy, 2013 ; Mons, 2008). Le rapport Eurydice (2009) confirme que dans les années 1980-1990, les systèmes européens ont vu se développer l'emploi systématique d'évaluations standardisées des acquis des élèves (nationales ou régionales) pour accompagner les changements en faveur de la décentralisation et de l'autonomie des établissements. Toutefois, ces dispositifs sont établis de manières différentes selon les contextes, tant dans leur fréquence que dans leur instrumentation (Maroy, 2005b). Ces différences dépendent

⁸ Les différentes fonctions politiques et pédagogiques de l'évaluation externe des acquis des élèves sont l'objet du point 5 (p.33-48).

notamment de l'importance donnée aux résultats des évaluations externes dans les politiques éducatives (politiques d'*accountability*).

3.4.4. *Les politiques d'accountability : les résultats de l'évaluation externe des acquis des élèves comme outils d'action publique*

Dans le modèle de l'Etat évaluateur, les résultats des élèves aux évaluations standardisées sont des outils d'action publique. Ils permettent au pouvoir central d'agir sur les acteurs de terrain qui lui sont redevables de leur autonomie. On appelle ce mécanisme « politique d'*accountability* » qui est défini par Mons & Durpiez (2010) comme « le couplage de mesures de la production du système éducatif (le plus souvent les acquis des élèves) et de responsabilisation des établissements scolaire, eu égard aux résultats obtenus par les élèves ». Pour Broadfoot (2000), l'*accountability* est un processus en deux étapes : 1) identification des performances du système par rapport aux objectifs fixés ; 2) actions des institutions par rapport aux écarts observés.

Le terme anglophone *accountability* est difficilement traduisible en français. Les termes francophones « reddition de comptes » ou « imputabilité » n'en constituent qu'une traduction partielle qui ne recouvre pas forcément ses deux étapes (prise d'informations et actions) (Maroy, 2013). Selon Mons (2009), les politiques d'*accountability* ont généralement comme but l'amélioration de la qualité de l'enseignement en jouant sur plusieurs aspects : diminution des inégalités scolaires, augmentation de la motivation et amélioration des performances des élèves, mobilisation plus importante des enseignants, meilleure information des autorités locales et nationales. De plus, pour cette auteure, ces politiques sont généralement « déclenchées » par l'utilisation d'évaluations externes standardisées des acquis des élèves. Quatre composantes définissent les politiques d'*accountability* (Harris & Herrington, 2006) :

- a) Formulation de standards ;
- b) **Evaluation externe** pour contrôler l'atteinte des standards ;
- c) Publication des résultats de ces évaluations externes ;
- d) Sanctions ou récompenses pour les écoles et/ou les enseignants.

Avec l'introduction des politiques d'*accountability*, les résultats des évaluations standardisées des acquis des élèves prennent encore davantage d'importance dans la gestion des systèmes éducatifs (Linn, 2000). Pourtant, pour Carnoy & Loeb (2002), l'évaluation des résultats des élèves et l'*accountability* n'alliaient auparavant pas forcément de pair. L'évaluation servait surtout à sélectionner les élèves pour des voies d'études différentes, l'*accountability* était surtout basée sur le contrôle des parents et des autorités locales sur la qualité de l'école. Aujourd'hui, les écoles doivent rendre des

comptes au niveau de l'Etat, mais ce qui est nouveau c'est que l'indicateur de la qualité est la performance des élèves à un test standardisé.

Les politiques d'*accountability* ne sont pas nouvelles, elles apparaissent dans les années soixante aux USA (Maroy, 2013 ; Meuret, 2012). Deux facteurs sont à leur origine : 1) la critique des milieux économiques du niveau scolaire faible des élèves américains (notamment par le rapport *A Nation At Risk*) ; 2) des préoccupations quant à l'équité des chances pour les différents élèves issus de groupes sociaux différents (confirmée par la loi fédérale *No Child Left Behind*). Après les USA, qui font office de pionniers dans le domaine, les politiques d'*accountability* se sont se sont répandues et ont gagné en importance également en Europe (Meuret, 2012). Dans les années 70 et 80, les Etats réforment leurs systèmes et proposent des indicateurs utiles aux établissements, rendus autonomes, pour analyser leurs performances. Cependant, dans les années 90, l'échec de ces réformes encourage les politiques à ajouter des mesures aux évaluations externes afin d' « [...] inciter ou contraindre les écoles et les enseignants à développer leurs capacités éducatives » (Maroy, 2013). Ces « sanctions » ne visent généralement pas à réduire les ressources allouées aux écoles mais à diminuer leur autonomie. Aux USA, la compétence pour ces sanctions est laissée aux Etats qui en choisissent leur degré (Meuret, 2012). Carnoy & Loeb (2002) tracent dans leur article un panorama des Etats américains et de leurs politiques d'*accountability*.

Les politiques d'*accountability* ont toutefois des caractéristiques très diverses, liées aux contextes dans lesquels elles s'inscrivent. Elles peuvent être différenciées et classées selon le niveau d'enjeux des sanctions et des récompenses pour les écoles, les enseignants et les élèves. Deux types d'*accountability* sont généralement distingués (Broadfoot, 2000 ; Carnoy & Loeb, 2002 ; Mons, 2009 ; Maroy, 2013) :

- *Accountability* « dure ». Ce type de politique est caractérisé par l'utilisation de tests standardisés à forts enjeux (*high stakes testing*) dont les résultats sont accompagnés de sanctions ou de récompenses qui ont des conséquences importantes pour les divers acteurs :
 - Élèves. Les résultats aux évaluations externes standardisées conditionnent leur carrière scolaire (certification d'un niveau d'étude, passage au niveau supérieur ou redoublement, sélection dans des voies d'études)
 - Enseignants. Ce type de politique vise la responsabilisation des enseignants par la publication et le classement des résultats des écoles aux évaluations externes standardisées, mais peuvent aussi prévoir des incitants plus directs (primes financières en cas de bons résultats, suivi par un supérieur hiérarchique, mutation ou licenciement en cas de mauvais résultats).

- Établissements. Les résultats aux évaluations externes standardisées peuvent créer la bonne ou la mauvaise réputation des établissements, mais aussi entraîner des mesures (diminution de l'autonomie des établissements placés sous la tutelle d'un organe étatique supérieur, fermeture de l'établissement, remplacement du personnel ou de la direction, primes collectives).

Le modèle d'*accountability* dure est plutôt le fait des systèmes scolaires anglo-saxons (USA, Angleterre, Pays de Galles, Australie). L'usage d'incitants est surtout dû à leurs mécanismes de libre-choix des écoles (Maroy, 2013). Toutefois, selon Normand & Derouet (2011), les pays anglo-saxons semblent sur le point d'adapter leur système vers un modèle d'« obligation de résultats intelligente », ceci en s'orientant davantage vers une auto-évaluation des établissements et en diminuant l'importance des incitants (comme en Ecosse). Cette tendance est également observée dans le rapport Eurydice (2009) et serait surtout due aux résistances des enseignants.

- *Accountability* « douce ». Ce type de politique est caractérisé par l'utilisation de tests standardisés à faibles enjeux (*low stakes testing*) dont les résultats n'entraînent pas ou peu de sanctions ou de récompenses pour les divers acteurs, mais servent surtout à informer les décideurs dans le but d'ajuster leurs actions. Le modèle de l'*accountability* « douce » est plutôt européen continental où d'ailleurs le terme *accountability* n'est que peu usité au bénéfice de ceux de « pilotage », « évaluation des politiques éducatives » ou « obligation de résultats » (Maroy, 2013). Mons & Durpiez (2010) distinguent deux modèles principaux d'*accountability* douce :

a) Le modèle développé par Klieme et al. (2004) influence les politiques éducatives de pays tels que l'Allemagne, la Suisse et l'Autriche. Il comprend la définition de « standards de formation »⁹ dont le degré d'atteinte est vérifié par des évaluations externes standardisées à faibles enjeux, le but étant d'améliorer la qualité de l'école et de l'enseignement en informant et orientant les divers acteurs :

- Élèves. Les résultats des évaluations externes standardisées permettent une meilleure information sur les priorités à suivre dans leurs apprentissages et sont un moyen de communication entre parents et enseignants. Ils n'ont par contre en aucun cas d'impact sur la carrière scolaire des élèves (promotion ou certification).

⁹ « Les standards de formation se réfèrent aux objectifs de l'éducation, poursuivis par l'apprentissage scolaire, et transposent ceux-ci en exigences concrètes » (Klieme et al., 2004, p. 17).

- Enseignants. Les résultats des évaluations externes standardisées permettent d'orienter le travail des enseignants et vise leur « responsabilisation » quant aux résultats de leurs élèves. Par contre, ces résultats ne sont pas dévoilés publiquement, ils restent le fait du corps enseignant et de leur hiérarchie.
- b) Le modèle développé par Thélot (2002), dit de l'« effet miroir », est une politique instaurée en France. L'évaluation externe standardisée doit permettre aux acteurs de se confronter aux résultats de leurs actions dans le but d'une amélioration, sans toutefois leur apporter de « schémas explicatifs ». Il n'y a donc qu'une sanction symbolique.

Toutefois, certains chercheurs, sur la base de travaux européens, observent actuellement des évolutions dans les politiques d'accountability des pays européens. Les politiques de « pilotage rationnel » des systèmes éducatifs de certains pays (France, Belgique et Suisse, mais aussi du Québec) sont en mutation. Elles semblent se « durcissent », se rapprochant des politiques d'accountability anglo-saxonnes sans toutefois prévoir de véritables sanctions matérielles (Demailly, 2001 ; Maroy, 2013 ; Mons, 2008).

Cette tendance au changement des modes de régulation conçue en termes d'une obligation morale de penser aux résultats, doublée d'une incitation cognitive et morale à la réflexivité, à partir des connaissances qu'on a des résultats de son action pourraient bien être renforcées dans un sens plus « impératif » et contraignant à l'avenir. Ainsi, certains Etats européens tendent à aller dans cette direction en couplant leurs politiques d'évaluation ou de pilotage, avec des « mécanismes de reddition de comptes » qui vont au-delà de la seule obligation de réflexivité demandée aux enseignants ou aux directions [...] (Maroy, 2013, p.16).

3.4.5. Les évaluations externes internationales : modèles de gouvernance par les résultats

Les systèmes éducatifs et leurs politiques ont évolué dans les différents contextes mondiaux vers un modèle semblable, celui de l'Etat évaluateur (standards de formation, évaluation standardisée, pilotage du système), tout en gardant des spécificités propres (différences dans les statuts et l'utilisation de ces outils). Outre les changements sociaux et économiques exposés auparavant, ces évolutions sont également liées à l'influence grandissante d'organisation inter ou transnationales telles que l'OCDE, l'UNESCO ou la Banque Mondiale, qui font la promotion de ce modèle de régulation comprenant l'usage de l'évaluation externe des acquis des élèves (Buisson-Fenet & Pons, 2011 ; Mons, 2007b).

En effet, la globalisation du modèle de l'Etat évaluateur permet des comparaisons internationales quant à la qualité (les résultats) des systèmes éducatifs (Maroy, 2012). Selon le rapport Eurydice (2009), les années 2000 ont été marquées par le développement d'enquêtes internationales telles que (parmi les plus connues) PISA (*Programme for International Student Assessment*), TIMSS (*Trends in International Mathematics and Science Study*) ou PIRLS (*Progress in international Reading Literacy*

Study). Les enquêtes internationales sont devenues une référence incontournable, voire un modèle, pour les politiques éducatives. Leurs résultats apparaissent dans les médias et sont un élément essentiel dans les prises de positions politiques sur l'école. (Bottani, 2004 ; Lafontaine, Soussi, & Nidegger, 2009 ; Normand, 2005). Les décideurs leur voient beaucoup d'avantages. Premièrement, les enquêtes internationales ont démocratisé les débats sur la qualité des systèmes éducatifs. Deuxièmement, on n'y prend plus seulement en compte les ressources ou le fonctionnement de l'enseignement comme unique indicateur, on s'intéresse plutôt à ses résultats. Aussi, son caractère quantitatif permet une observation à large échelle et une comparaison mondiale des systèmes éducatifs et des pratiques pédagogiques. Troisièmement, elles permettent d'identifier l'influence de variables individuelles, scolaires et organisationnelles sur les performances (Mons, 2008).

Les enquêtes internationales évaluant les acquis des élèves ont eu pour conséquence d'accélérer l'harmonisation des contenus au niveau international et national et de leur évaluation. Bottani (2004) donne l'enquête PISA comme exemple de ces effets en Suisse et en Allemagne notamment. Les enquêtes internationales ont inspiré les décideurs qui, suivant leur modèle (évaluation à large échelle des acquis des élèves), veulent leur propre outil d'information (Eurydice, 2009). Selon Eurydice (2009), les résultats des enquêtes internationales ont poussé bon nombre de pays à introduire ou à réintroduire des évaluations standardisées nationales dans le but de mesurer et d'améliorer la qualité de leur système éducatif (Hongrie, Autriche, Lituanie, Islande) ou de développer encore davantage les tests standardisés pré-existants (Danemark, Luxembourg). Les comparaisons internationales rendues possible par de ces enquêtes de grande envergure ont ainsi encore développé davantage l'utilisation d'épreuves standardisées externes.

3.5. Les critiques du modèle de l'Etat évaluateur

De nombreux chercheurs mettent en doute le modèle de l'Etat évaluateur (ou de l'obligation de résultats). Ils s'interrogent surtout sur ses capacités à améliorer la qualité de l'école. Ils lui adressent ces différentes critiques :

- a) *Des critiques au niveau méthodologique.* Tout d'abord, le fait de formuler des objectifs clairs et quantifiables (standards) peut être perçu comme utopique et difficilement réalisable au niveau méthodologique. Ensuite, c'est surtout le fait qu'on évalue la globalité du systèmes éducatifs par un seul indicateur (les acquis des élèves) qui ne fait pas l'unanimité (Demailly, 2001 ; Duru-Bellat & Jarousse, 2001). L'aspect exclusivement quantitatif de ces tests semble peu pertinent et devrait être complété par d'autres informations plus qualitatives (Bélaïr, 2005; Blais, 2005 ; Demailly, 2001). Meuret (2012) avance la « loi de Campbell », concept économique qui met en garde contre l'utilisation de résultats quantitatifs (ici les résultats des

élèves) comme unique élément de prise de décisions, ce qui peut résulter sur des effets pervers :

Plus un indicateur quantitatif est utilisé pour prendre des décisions, plus il sera sujet à des pressions corruptrices et plus il sera susceptible d'altérer et de corrompre le processus social qu'il est censé contrôler (Campbell, 1976, cité par Meuret, 2012).

En effet, les apprentissages de l'élève (ou sa performance) dépendent de bien plus de facteurs que des simples réformes politiques. Aussi, le principe d'uniformité est critiqué. La variabilité des élèves et des enseignants n'est que peu prise en compte par cette forme d'évaluation des politiques (Perrenoud, 1998 ; Duru-Bellat & Jarousse, 2001).

La démarche d'évaluation par les résultats fait comme si tous les enseignants partageaient des finalités identiques, celles mises en avant par les textes. Il est dans doute vrai que face à des textes, des directives, des programmes souvent ambigus, les enseignants s'adaptent, voire pratiquent une résistance passive, quand ils ne partagent pas la philosophie. Il serait alors simpliste de prétendre évaluer les politiques affichées dans les textes dans des contextes concrets où les objectifs initiaux n'ont plus forcément cours, tant ils ont été réinterprétés, détournés... ou ignorés (Duru-Bellat & Jarousse, 2001, p. 104).

Le recours aux acquis des élèves comme unique indicateur de la qualité du système relève plus de la commodité. Blais (2005) explique cet attrait exclusif par des coûts moindres et une mise en place rapide.

- b) *Des critiques quant à la réception de ces politiques par les acteurs de terrain.* L'évaluation des politiques éducatives comme processus *top-down* ne serait que peu efficace, car elle est davantage perçue par les acteurs de terrain comme un contrôle qui remet en cause leur autonomie. Leur portée sur le terrain resterait alors faible (Duru-Bellat & Jarousse, 2001 ; Suchaut, 1998). Elle serait même la cause d'une certaine « déprofessionnalisation » des enseignants (Mons & Durpiez, 2010 ; Maroy & Cattonar, 2002). En effet, ces politiques engagent un glissement de la responsabilité de l'Etat vers celle des écoles et/ou des enseignants.

« [...] les évaluations de l'efficacité des écoles sont parfois utilisées de manière très offensive dans le cadre de véritables campagnes de dénonciation qui condamnent les écoles plus qu'elles ne les aident » (Duru-Bellat & Jarousse, 2001, p. 107).

- c) *Des critiques quant à l'efficacité de ces politiques.* Pour certains chercheurs, certaines politiques d'*accountability* ne seraient pas une avancée pour les systèmes éducatifs. Des recherches ont démontré qu'elles ne contribuent ni à améliorer la performance des élèves (Linn, 2000 ; McDonnell, 1994), ni à réduire les inégalités (Mons, 2013 ; Berliner & Nichols, 2007). Par contre, d'autres chercheurs identifient des liens entre politiques d'*accountability*

dures et amélioration des performances (Hanushek & Raymond, 2004) et/ou réduction des inégalités (Carnoy & Loeb, 2002).

- d) *Des critiques concernant les enquêtes internationales.* Certains chercheurs émettent des critiques au sujet des enquêtes internationales sur le plan méthodologique mais aussi sur celui de leur exploitation par les médias et la politique. Premièrement, le caractère quantitatif de ces enquêtes ne permet pas d'observer l'élève dans son ensemble. De plus, la fiabilité des résultats est parfois remise en cause. Des biais sont observés dans le choix des items et l'échantillonnage des tests, dans la traduction des items et les biais culturels. Aussi, la comparaison entre les systèmes éducatifs paraît parfois peu pertinentes, en lien avec leurs caractéristiques très diverses. Deuxièmement, les chercheurs mettent en garde contre l'exploitation parfois erronée ou malintentionnée de ces enquêtes : les résultats et les palmarès sont parfois trop simplifiés, ne prenant pas en compte des marges de fiabilité statistique. Finalement, le coût de ces enquêtes à large spectre est souvent relevé comme trop élevé, mobilisant les budgets de recherche des pays. (Mons, 2007a, 2008 ; Monseur & Demeuse, 2004 ; Normand, 2005). Aussi, ces épreuves internationales, dû à leur caractère trop macro, n'affectent que très peu les enseignants dans leur classe. Ceux-ci n'y prêtent que très peu d'attention (Lafontaine et al., 2009).
- e) *Des critiques quant aux effets indésirables et néfastes de ces politiques.* Finalement, ces politiques, par le biais des évaluations externes, ont des effets indésirables et néfastes sur les acteurs de terrains¹⁰.

¹⁰ Les effets de l'évaluation externe sont l'objet du point 6 (p.49-74).

4. Développement et usage de l'évaluation externe dans différents contextes

Après avoir présenté les différents usages de l'évaluation externe des acquis des élèves et les raisons de son développement, la présente partie a pour objet d'en donner quelques illustrations dans différents contextes nord-américains et européens. Le cas de la Suisse est présenté dans une partie indépendante (chapitre 2). Pour des informations sur d'autres pays, nous renvoyons le lecteur aux synthèses de Soussi & Nidegger (2012) ou Looney (2009).

4.1. L'évaluation externe en Angleterre

Le système éducatif anglais fait office de précurseur dans l'usage des évaluations standardisées en tant qu'outil central de leur système éducatif (intégré dans un système de reddition de compte). En Europe, l'Angleterre est certainement le pays où la culture de l'évaluation externe est la plus forte. Nous nous référons aux travaux de Broadfoot (2000) et Dupriez (2005) pour en faire une brève description.

Héritage de son passé féodal, l'Angleterre connaissait il y a quelques décennies encore un système éducatif décentralisé. Toutefois, l'Etat central jouait déjà un rôle de contrôle. Il existait déjà une tradition ancienne de l'évaluation externe dans le but de certification des élèves et de contrôle de la qualité de l'enseignement (Inspecteurs de sa Majesté, examens externes). Ces procédures comprenaient d'ailleurs déjà des sanctions pour les établissements (subventions) et les élèves (carrière scolaire). Toutefois, les enseignants jouissaient d'une grande liberté d'action.

Dans les années 1970, on observe une inquiétude grandissante au sujet du manque d'information sur le système et sur le niveau de performance des élèves. Ce qui se traduit en 1988, par une réforme éducative qui renforça le rôle de l'Etat central. Un curriculum national et un programme d'évaluation à différents degrés de la scolarité furent introduits. Ces évaluations externes nationales (qui vinrent s'ajouter à celles locales) servent encore aujourd'hui à informer les élèves (et leurs parents) sur leur performance scolaire, mais également les autorités sur la qualité des enseignements et des établissements. Ces mesures sont accompagnées par des évaluations externes plus qualitatives (inspections des établissements). Une entité a été créée pour mener ces investigations, l'*Office for standards in Education (OFSTED)*. Cette procédure n'est pas sans conséquences pour les enseignants et les établissements, le système éducatif anglais étant inscrit dans une politique d'*accountability*. Premièrement, les résultats de ces évaluations externes sont remis au conseil d'administration de chaque établissement qui prend des mesures pour améliorer ses résultats (plan d'action, formulation d'objectifs précis). Deuxièmement, les parents, qui ont le libre choix de l'école pour leurs enfants, y ont accès, un résumé du rapport étant spécialement conçu pour eux et publié sur Internet.

Troisièmement, un classement des établissements est établi (*league tables*). Finalement, il semble que ces politiques ont permis d'améliorer le niveau moyen de l'éducation en Angleterre. Toutefois, les écarts entre établissements se creusent et le système perd en équité pour les élèves également. En outre, il est légitime de s'interroger sur la « normalisation » de ces dispositifs, les établissements et enseignants recevant des instructions claires, limitant leur degré de liberté pédagogique.

4.2. L'évaluation externe aux Etats-Unis

Le système éducatif américain inspiré à la base du modèle anglais, est considéré aujourd'hui comme celui qui fait la plus grande place à l'usage des évaluations externes. Nous présentons ce développement en nous référant aux travaux de Malet (2011), Meuret (2012) et Normand (2005).

Les épreuves externes sont partie intégrantes des politiques d'*accountability* instruites dès les années 1980 par des réformes du système éducatif américain. Ces mutations organisationnelles prennent leur origine dans le constat du faible niveau des élèves américains en comparaison avec ceux d'autres pays. Le rapport *A Nation at Risk* (1983) constitue un premier tournant dans la mutation des politiques éducatives américaines. Ce rapport politique remet en cause l'autonomie laissée aux établissements, les compétences des enseignants et le choix des contenus enseignés. On prône alors un retour aux fondements (*back to basics*) et un modèle managérial basé sur les résultats et la compétition entre écoles. Dans la plupart des Etats, les contenus sont alors standardisés et on prévoit un pilotage du système basé sur les résultats des élèves (mesuré par des tests standardisés). Des systèmes d'incitations (récompenses ou sanctions) liées aux résultats de ces tests sont également introduits, mais reste au libre choix des Etats (souverains en matière d'éducation). Entre 1990 et 2002, la moitié des Etats adoptent leurs propres procédures d'*accountability* plus ou moins dures (publication des résultats des établissements avec ou sans incitants).

Puis, inspirée des évaluations nationales et internationales (qui montrent notamment le niveau faible des élèves américains) une loi fédérale appelée *No Child Left Behind (NCLB)* est promulguée en 2002. L'Etat fédéral impose alors à tous les Etats d'introduire des procédures d'*accountability* (les modalités étant propres à chacun), sous peine de ne pas recevoir les financements fédéraux pour leurs programmes éducatifs. Les Etats sont de ce fait tenus de standardiser leurs contenus d'enseignement, d'évaluer leur atteinte par des tests standardisés (*high stakes tests*) et de prévoir l'usage d'incitants basés sur les résultats des tests. Les Etats doivent alors fixer des compétences à atteindre par les élèves (lecture, maths, sciences) qui sont testées régulièrement (chaque année de 8 à 13 ans, puis une fois en *High School*). Ils doivent faire en sorte que tous les élèves atteignent un certain niveau de compétences et définissent des objectifs pour chaque école. Les établissements doivent également rédiger un rapport sur les performances des élèves selon certains facteurs (race, ethnicité, niveau socio-économique,

degré de performance, niveau de langue). Si ceux-ci ne parviennent pas à atteindre leurs objectifs annuels, des mesures sont prises par l'Etat (cours supplémentaires aux élèves en difficulté, possibilité donnée aux parents de changer d'école, changement d'enseignant, restructuration de l'établissement, prise sous tutelle de l'école par l'état). De plus, dans le but de prouver leurs efforts, les Etats doivent se soumettre à un test national le *Nation's Report Card* appelé autrefois le *National Assessment of Educational Progress (NAEP)*. Enfin, en 2009, le programme fédéral *Race to the top* a relancé encore l'exigence faite aux Etats de mettre en place ces politiques d'*accountability* (Mons, 2009).

Les politiques d'*accountability* dures imposées aux Etats et les tests qui y sont liés partagent les chercheurs américains. L'efficacité de telles politiques est remise en cause. Elles ne permettraient en effet pas de réduire les inégalités, ni d'améliorer le niveau d'instruction des élèves.

4.3. L'évaluation externe en France

La France, pays dont le système scolaire est fortement centralisé, est un modèle que l'on peut considérer comme « non-linéaire » dans l'usage de l'évaluation externe des acquis des élèves. En effet, son développement s'est réalisé de manière « chaotique » dû aux nombreuses oppositions politiques et professionnelles. Nous décrivons cette évolution en synthétisant l'état des lieux de Mons & Pons (2006).

Au milieu des années 1970, deux facteurs sont à l'origine des premières évaluations externes en France : 1) l'ouverture de la France aux enquêtes internationales ; 2) une volonté d'évaluer les réformes de politique éducative (notamment celle du collège unique). Naît alors la DEPP (Direction de la prospective et de la performance), organe ministériel chargé de mener des évaluations externes de manière systématique dès les années 80. Ces dispositifs sont définis comme diagnostiques, servant surtout à aider les enseignants dans leur classe. Justement, ces épreuves sont mal accueillies par les enseignants qui se méfient de cette standardisation. On craint alors une évaluation des enseignants au travers de celle de leurs élèves. Dans les années 90, les évaluations nationales restent diagnostiques et quelques évaluations bilan se développent dans le but de juger les performances des élèves à certains niveaux. D'ailleurs, les fonctions de ces différents tests sont mélangées par les politiques et les enseignants. Au début des années 2000, un changement de pouvoir fait ralentir le développement des épreuves nationales. Cependant, en réponse, les autorités locales (inspecteurs et académies) créent leurs propres outils d'évaluation externe. Enfin, dans les années 2000, les épreuves diagnostiques et bilans sont relancées, sous l'influence des résultats de l'enquête PISA et de décisions politiques pour la rentabilisation des actions publiques.

Le système français d'évaluation des acquis des élèves se résume donc à l'usage important mais quelque peu « disparate » de deux types d'évaluation externe : diagnostique (évaluation de masse,

servant à informer les enseignants sur les difficultés de leurs élèves en début d'année) et bilan (évaluation par échantillon, servant à informer les décideurs sur l'évolution des politiques éducatives). Ces dispositifs comportent des enjeux limités, le pilotage du système restant encore discret, la France n'ayant pas réellement de standards de formation (mais un « socle commun » mais sans mesure d'évaluation qui l'accompagne). Enfin, l'épreuve du baccalauréat reste une épreuve certificative (qui est standardisée) qui tient beaucoup d'importance dans le système français et qui fixe notamment le niveau attendu des élèves.

4.4. L'évaluation externe en Allemagne

Le système éducatif allemand connaît également le développement de l'évaluation externe (*externe evaluation*). Nous faisons ici l'état des lieux, en nous basant sur le livre de Döbert et Dederling (2008). Dans les années 1990 à 2000, l'Allemagne ou plutôt ses *Länder* (souverains en matière d'éducation depuis 2005) ont introduit de nouveaux outils dans le but de contrôler/piloter la qualité de l'école. Ce développement s'est fait surtout par l'influence des enquêtes internationales (PISA, PIRLS...). On s'interroge alors sur le niveau du système. Les administrations veulent se doter d'outils capables de mesurer la qualité de l'enseignement. Tous les Länder adoptent leur propre système de pilotage. Ceux-ci sont disparates, intégrant avec une importance différente des outils semblables : formulation de standards, épreuves externes, inspections des écoles, pilotage. Ces dispositifs sont décrits pour chaque région dans l'ouvrage cité. Plusieurs catégories se distinguent : *externe Evaluation, Fremdevaluation, Qualitätsanalyse, Schulevisitation*. La différence entre régions étant marquée, aucune comparaison nationale n'est possible.

Il est à noter que dans les procédures mises en place, les épreuves externes (*standardisierte Tests*) servent d'indicateurs au service d'évaluations plus qualitatives. Elles n'ont pas une place aussi importante que dans les autres contextes décrits. D'ailleurs, l'Allemagne n'a introduit un test national qu'en 2009 (Eurydice, 2009). Celui-ci n'est utilisé que pour la certification des élèves à la fin de la scolarité obligatoire. Il ne sert pas au pilotage du système et ses résultats ne font office d'aucune publication.

4.5. L'évaluation externe au Canada

Le système éducatif canadien est fortement influencé par les systèmes anglais et américains et s'est aujourd'hui orienté vers une politique d'évaluation standardisée. Nous en faisons un petit état des lieux en nous référant aux travaux de Bélair (2005) et Mons & Pons (2006).

Bien qu'au Canada la tradition des tests provinciaux ou à large échelle pour la certification des élèves soit plus ancienne, le développement des épreuves externes tient son origine dans une crise de l'éducation publique. Dans les années 1980 (comme aux USA), la qualité de l'école fut remise en cause par des attaques politiques. Un système de redevabilité basé sur les résultats a donc été graduellement mis en place dans le but d'améliorer la qualité de l'éducation. Aujourd'hui, les provinces de ce système fédéral ont mis en place leur propre dispositif d'évaluation standardisée. Les tests ont deux objectifs principaux : évaluer les apprentissages des élèves mais aussi mesurer les performances du système (et des écoles). En effet, les résultats des épreuves sont utilisés pour des comparaisons entre écoles et entre provinces. Toutefois, leurs finalités sont diverses selon les provinces : certification des élèves, diagnostic, évaluation des établissements... Les résultats sont, généralement, publiés et créent une certaine compétition entre écoles. Cependant, les épreuves se différencient de celles des USA, ne comportant explicitement pas de forts enjeux (il ne s'agit pas de *high stakes tests*), aucune sanction n'étant prévu en cas de mauvais résultats. Finalement, l'impact de ces tests soulève l'interrogation de nombreux chercheurs.

4.6. L'évaluation externe en Belgique francophone

L'évaluation standardisée s'est développée dès les années 1990 en Communauté francophone de Belgique (CFB). Historiquement, son système éducatif décentralisé laisse une grande marge de manœuvre aux acteurs locaux (certification des élèves, moyens pédagogiques) et aucun processus central n'était prévu. Toutefois, certains rapports internationaux (notamment de l'OCDE) ont mis en lumière certaines carences de ce système : redoublement fréquent des élèves et forte disparité de résultats entre établissements. Ces indicateurs ont conduit le Ministère de l'Éducation à mettre en place des outils centralisés pour lutter contre ces problèmes, sous forme d'évaluation externe des acquis des élèves (Lafontaine, 2000).

Deux périodes d'implantation de ces dispositifs externes peuvent être différenciées (Lafontaine, 2012). La première période voit la création d'évaluations externes diagnostiques qui devaient servir à l'enseignant dans sa classe, sans finalité certificative. Par la suite, un texte de loi majeur va fournir aux autorités de CFB les moyens de mener un véritable pilotage centralisé : le « décret Missions » de 1997. Ce texte définit les buts du système éducatif, officialise un référentiel de compétences central qui vient s'ajouter aux programmes d'études des différentes entités locales et prévoit des outils d'évaluation. Ces évaluations devaient servir à sortir les enseignants de leur « isolement » en leur donnant des moyens de comparer les résultats de leurs élèves et des pistes didactiques. Entre 2002 et 2009, de nouveaux textes juridiques font accélérer l'usage de ces dispositifs. L'objectif est cette fois de fournir des informations aux acteurs politiques de l'éducation par de nouvelles épreuves externes

standardisées, cette fois certificatives (en fin de primaire, en début et en secondaire). Les épreuves non certificatives sont aussi renforcées.

Finalement, en moins de deux décennies, le système éducatif de la CFB a vu naître une culture de l'évaluation qui s'est développé rapidement. L'évaluation externe standardisée y prend aujourd'hui une grande importance (surtout par le nombre d'épreuves), mais le pilotage y reste « bienveillant » : les évaluations non certificatives ne servant pas à l'évaluation des établissements et des enseignants. (Lafontaine, 2012).

Toutefois, de vives interrogations émanent de la communauté scientifique belge. Tout d'abord, les autorités n'ont pas été jusqu'au bout de leur politique de pilotage, ne mettant notamment que peu d'épreuve à disposition des enseignants ou la finalité de certaines étant double (diagnostique et bilan). Ceci entraînerait d'ailleurs l'incompréhension des enseignants et leur résistance à cette politique de l'évaluation (Dupriez, 2004; Rey, 2005).

5. Fonctions de l'évaluation externe des acquis des élèves

Résultant de l'évolution des systèmes éducatifs (exposée dans le point précédent), l'usage de l'évaluation externe des acquis des élèves s'est développé de manière importante lors des dernières décennies. Elle remplit aujourd'hui de multiples fonctions à tous les « niveaux » des systèmes éducatifs, qui sont d'ailleurs interdépendants (Mottier Lopez, 2009) :

- *Au niveau micro* : niveau de la classe (enseignement-apprentissage)
- *Au niveau méso* : niveau de la gestion locale (établissement ou région)
- *Au niveau macro* : niveau du système éducatif et des politiques éducatives

Dans la littérature, les fonctions de l'évaluation externe sont catégorisées de diverses manières. Dans cette partie nous exposons premièrement ces différents classements avant d'en faire une proposition synthétique en deux catégories : ses fonctions « pédagogiques » et ses fonctions « politiques ». Deuxièmement, nous détaillons les fonctions propres à ces deux catégories.

5.1. Les fonctions multiples de l'évaluation externe

De nombreux travaux présentent une catégorisation des fonctions de l'évaluation externe. Nous en proposons quelques-unes, avant d'en faire une proposition synthétique.

- Pour Weiss (2002), l'évaluation, aussi bien interne qu'externe, remplit de multiples fonctions aussi bien pour l'élève et ses parents, que pour l'enseignant, l'établissement ou le système de formation (tableau 2). Pour les deux approches d'évaluation, l'auteur distingue deux « logiques » : la logique de l'apprenant et la logique de l'organisation.

TABLEAU 2 – Les différentes logiques de l'évaluation (Weiss, 2002)

	Logique de l'apprenant	Logique de l'organisation
Évaluations externes	épreuves diagnostiques pour l'évaluation des apprentissages des élèves	évaluation bilan des systèmes de formation (cantons, pays)
	épreuves bilan pour la certification des apprentissages des élèves	évaluation pour la certification des organisations apprenantes
Évaluations internes	évaluation formative des apprentissages des élèves	évaluation régulatrice autogérée d'une organisation apprenante
	Évaluation bilan-sommative des apprentissages de élèves	

- La langue anglaise et les travaux de recherches anglophones distinguent deux logiques de l'évaluation. Le terme *assessment* caractérise la logique de l'apprenant et le terme *evaluation* la logique de l'organisation (Madaus & Kellaghan, 1992).
- Monseur & Demeuse (2005) se référant aux travaux de McDonnell (1994) relèvent également une double fonction de l'évaluation externe :

Les objectifs poursuivis par ces épreuves (externes) sont multiples et peuvent être décrits sur un continuum : d'un côté, le pôle informatif, l'objet est de rassembler des informations sur le système éducatif et d'aider à prendre des décisions relatives à l'instruction des élèves. De l'autre côté, le pôle régulation, il s'agit d'une perspective de contrôle des écoles, des enseignants et des élèves (p.493).

- Dans la figure 1, Wirthner & Ntamakiliro (2008) présentent également la position des épreuves externes au centre d'un « continuum » entre évaluation des apprentissage des élèves et évaluation du système scolaire. Pour ces auteurs, dans chaque contexte (pays, région,...), les épreuves externes peuvent se rapprocher plus ou moins de l'une ou l'autre finalité. Toutefois, ces deux finalités sont aujourd'hui toujours présentes.

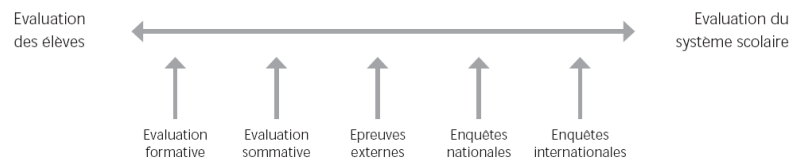


FIGURE 1 – Epreuves externes sur un continuum (Wirthner et Ntamakiliro, 2008, p.92)

- Dans le rapport Eurydice (2009), on distingue trois types d'évaluation standardisée nationales, différenciées selon les principaux objectifs poursuivis. Deux de ces types peuvent être assimilés à un rôle pédagogique (prise de décision sur le parcours des élèves ; identification des besoins d'apprentissage individuels des élèves) et la troisième à un rôle politique (pilotage des établissements et/ou du système éducatif). La figure 2 présente leur distribution dans les pays européens.

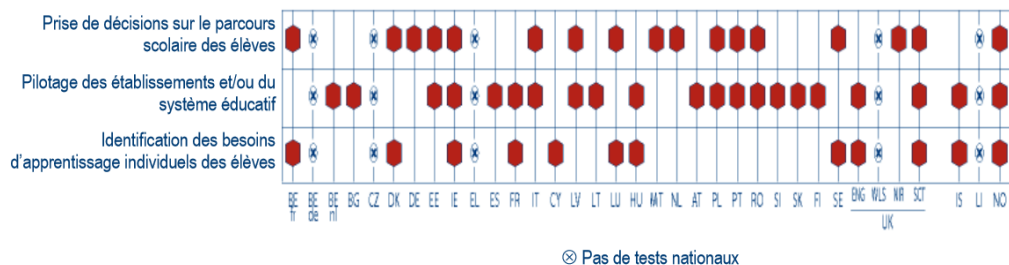


FIGURE 2 – Principaux objectifs des tests nationaux standardisés (niveaux CITE1 et 2, 2008-2009) (Eurydice, 2009, p.25)

Sur la base de ces diverses catégorisations, nous classons les différents rôles des évaluations externes des acquis des élèves en deux catégories :

- *Fonctions pédagogiques* (logique de l'apprenant – *assessment* – pôle informatif). Les épreuves externes complètent l'évaluation interne (évaluation formative ou sommative réalisée en classe) dans le but de favoriser ou de situer l'apprentissage de chaque élève ou d'un groupe d'élève.
- *Fonctions politiques* (logique de l'organisation – *evaluation* – pôle régulation). Les épreuves externes, se rapprochant des enquêtes nationales ou internationales, servent au pilotage du système éducatif.

Les épreuves externes remplissent toutefois souvent plusieurs fonctions simultanément, ce qui peut notamment créer des tensions¹¹. Dans les points suivants, nous présentons les diverses fonctions politiques puis pédagogiques de l'évaluation externe.

5.2. Les fonctions politiques de l'évaluation externe

L'évaluation externe des acquis des élèves est devenue un outil de politique éducative. Le rapport Eurydice (2009) montre que de plus en plus de pays (ou de régions) européens utilisent des tests standardisés dans un but de pilotage des établissements et/ou du système scolaire entier. Aujourd'hui, plus de la moitié des Etats européens utilisent ce type de tests nationaux avec des objectifs divers : comparaison et responsabilisation des établissements et/ou évaluation de la performance du système entier. Les résultats sont donc des indicateurs de la qualité de l'enseignement, des politiques et des pratiques éducatives. Ces tests sont généralement passés par des échantillons représentatifs d'élèves.

Toutes les fonctions politiques de l'évaluation externe se retrouvent dans le terme général de "pilotage"¹².

De façon proactive, ce pilotage (fondé sur les résultats des élèves) peut servir à réguler la mise en place de pratiques efficaces, à en suivre la progression et à en évaluer les effets. Dans un cas comme dans l'autre, l'évaluation des résultats, en particulier de l'apprentissage des élèves, est nécessaire au pilotage réussi du système, tant pour y découvrir des problèmes insoupçonnés que pour y mettre en place des correctifs appropriés (Laveault, 2009a, p. 177).

Mons (2009) différencie toutefois trois rôles différents : elle est à la fois un outil de contrôle, un outil de régulation et un outil de transfert des responsabilités. Nous détaillons ci-après ces trois finalités, en

¹¹ Ces tensions sont l'objet du point 5.4 (p.45-48).

¹² Ce terme est défini au point 3.2 (p.10).

distinguant la fonction d'outil *d'accountability* (comprise par Mons (2009) dans la fonction de contrôle).

5.2.1. L'évaluation externe, outil de contrôle des systèmes politiques

Pour Mons (2009), l'évaluation standardisée est un outil de contrôle qui produit des données quantifiables et comparables sur les systèmes éducatifs et sur les réformes engagées. La fonction de l'évaluation externe devrait donc être « un miroir réfléchissant des points brillants et des zones d'ombre du système d'enseignement » (Ntamakiliro, 2003, p.3). Elles renseignent les décideurs de l'atteinte ou non des objectifs fixés, mais donnent aussi diverses informations sur des variables utiles telles que la comparaison entre les élèves, les établissements ou les régions. Les résultats des épreuves des acquis des élèves, généralement sous forme de statistiques, servent à la vérification et au contrôle du degré d'efficacité, voire d'efficience, et d'équité du système à des niveaux divers, tel qu'un pays, une région, un canton ou un établissement (Ntamakiliro & Longchamps, 2008 ; Soussi & Niedegger, 2010). Pourtant, des observations et des analyses spécifiques ne sont que rarement mises en place (Broadfoot, 2000 ; Mons & Crahay, 2011).

[...] l'analyse fine des dispositifs mis en place et des usages qui sont faits des données recueillies par les épreuves standardisées montre une faible utilisation de ces données dans la formulation et/ou régulation des politiques éducatives. [...] ce sont davantage les acteurs opérationnels au niveau des établissements scolaires et des collectivités territoriales qui sont en charge de l'élaboration des politiques éducatives et qui sont la cible du mouvement de reddition de comptes (Mons & Crahay, 2011, p.90).

5.2.2. L'évaluation externe, outil de régulation des systèmes politiques

Pour McDonnell (1994), les épreuves externes sont très appréciées par les autorités, afin d'avoir un contrôle et un effet de pression sur ce qui se déroule dans les classes. Dans le champ des politiques éducatives, on parle alors de régulation¹³. Cette fonction consiste à orienter les conduites des acteurs de terrain vers des "règles", des normes d'action et de pratiques (Maroy & Dupriez, 2000). Pour Mons (2009), l'évaluation standardisée est un outil de « direction cognitive » qui permet aux autorités de réguler les actions et les pratiques sur le terrain¹⁴. Elle peut, par exemple, être le vecteur de lignes directrices dans les contenus prioritaires à enseigner et/ou dans les pratiques à adopter (Laveault, 2009 ; Mons, 2009 ; Mons & Crahay, 2011 ; Ntamakiliro & Tessaro, 2010).

¹³ Cette fonction est définie au point 3.2 (p.11).

¹⁴ L'effet de régulation des pratiques des enseignants est l'objet du point 6.2.3 (p.57-66).

Le « pilotage par l'évaluation » peut être vu comme une façon de réguler les comportements professionnels au niveau collectif comme individuel, en donnant une mesure la plus fiable possible de ce qui sépare les acteurs en situation des objectifs à atteindre (Felouzis & Hanhart, 2011, p. 13).

Toutefois, le terme de « régulation » peut être employé différemment selon les auteurs et leur approche. En effet, il est utilisé dans de nombreuses disciplines (par ex. économie, biologie, psychologie, sociologie, sciences de l'éducation, etc.). Dans une acceptation plutôt propre au champ des politiques éducatives, « régulation » signifie parfois « pilotage », « ajustement », « amélioration » ou tout simplement « changement ». Cependant, il est important de spécifier ce terme, grâce aux apports du champ de la régulation des apprentissages et/ou de l'enseignement (Allal & Mottier Lopez, 2005; Black & Wiliam, 2009). La régulation comprend quatre opérations (Allal, 2007, p. 8) :

- Fixer un but et orienter l'action vers celui-ci ;
- Contrôler la progression de l'action vers le but ;
- Assurer un retour sur l'action (un feedback, une rétroaction) ;
- Confirmer ou réorienter la trajectoire de l'action, et/ou redéfinir le but.

Le mécanisme de régulation peut être comparé à une boucle (*loop*) entre un état initial et un état final désiré. Mais pour être refermées, pour que l'écart (*gap*) entre ces deux états puisse être comblé, cette boucle doit comprendre quatre éléments essentiels : des objectifs à atteindre, des mesures initiales et finales, un feedback et une remédiation visant un ajustement (Laveault, 2007). Dans ce sens, l'évaluation externe peut être un outil de régulation car elle peut permettre de communiquer des attentes (ou objectifs) (par ex. pratiques attendues, contenus attendus, performances attendues, etc.) et de les mesurer. Néanmoins, tous les dispositifs d'évaluation externe n'offre pas toutes les composantes nécessaire à une régulation réussie : parfois le feedback (le plus souvent sous forme de résultats globaux) n'est pas suffisant ou suffisamment accompagné pour offrir une véritable remédiation. La boucle de rétroaction n'est, dans ce cas, pas refermée.

[...] information about the gap between actual and reference levels is considered as feedback only when it is used to alter the gap. If the information is simply recorded, passed to a third party who lacks either the knowledge or the power to change the outcome, or is too deeply coded (for example, as summary grade given by the teacher) to lead appropriate action, the control loop cannot be closed, and "dangling data" substituted for effective feedback (Sadler, 1989, p. 121).

5.2.3. L'évaluation externe, outil de transfert des responsabilités

Selon Mons (2009), l'évaluation standardisée est également un outil de transfert de responsabilités. Autrefois (dans les modèles bureaucratiques), c'est l'Etat qui était redevable de rendre des comptes quant aux réformes engagées, aujourd'hui (dans le modèle post-bureaucratique) cette responsabilité est

surtout mise à la charge des écoles (voire des enseignants) qui sont, en contrepartie, davantage autonomes. La responsabilisation des enseignants par l'évaluation externe est d'autant plus forte lorsque les résultats entraînent des sanctions ou des récompenses (Mons & Dupriez, 2010). Les évaluations externes servent alors au contrôle du travail des enseignants qui peut être individuelle ou collective (équipe pédagogique, établissement), mais aussi interne (auto-évaluation, évaluation par les pairs) et/ou externe (épreuves externes, inspection). Ce contrôle semble s'intensifier dans tous les pays européens autant dans leur fréquence que dans leurs enjeux. De plus, leurs formes traditionnelles d'inspection sur les processus sont remplacées petit à petit par des modalités d'évaluation basées sur les résultats. Pourtant, en Europe, rares sont les pays qui prévoient des récompenses aux enseignants pour la bonne performance de leurs élèves (Ecosse, Espagne, Hongrie, Lettonie, Roumanie) (Mons, 2008a).

Certains chercheurs en viennent même à se demander si les évaluations externes n'évaluent pas de fait plutôt le travail des enseignants que celui des élèves (Wirthner & Ntamakiliro, 2008). Toutefois, Mons et Pons (2006) s'interrogent : qui doit être tenu responsable de ces résultats ? Cette question a été débattue dans l'ouvrage dirigé par Lessard et Meirieu (2005) « L'obligation de résultats en éducation ». Meirieu (2005) qualifie d'« irresponsabilité sociale » le fait que les enseignants veulent se dédouaner complètement de toute responsabilité quant aux résultats de leurs élèves aux évaluations externes. Tardif (2005), lui, fait un plaidoyer pour une « responsabilité limitée » de l'enseignant quant aux résultats des élèves aux tests. La responsabilité serait partagée entre les différents enseignants qui ont eu la charge des mêmes élèves. Au contraire, pour Paul (2005), les enseignants ne peuvent pas être tenus comme responsables de ces résultats, au contraire des autorités. Enfin, pour Beillerot (2005), les nouvelles politiques de l'obligation de résultats font oublier la responsabilité des élèves eux-mêmes, tenus jadis comme principaux responsables. Il défend donc les enseignants qui sont selon lui responsable de moyens et non de résultats.

5.2.4. L'évaluation externe, outil d'*accountability*

Selon les contextes politiques, les résultats produits par l'évaluation externe des acquis des élèves permettent la mise en place de politiques d'*accountability*. Elles rendent les établissements et les enseignants responsables des résultats de manière interne envers l'institution (contrôle des écoles et des enseignants par les autorités scolaires) et de manière externe envers la société (informations sur la qualité du système et/ou des écoles et des enseignants) (Mons, 2009). Les épreuves externes ont un statut différent selon les formes d'*accountability*, catégorisées de « dures ou douces »¹⁵ :

¹⁵ Les différentes politiques d'*accountability* sont décrites au point 3.4.4 (p.20-23).

- Dans les pays anglo-saxons, où les politiques d'*accountability* sont dites « dures » (notamment dues à la publication des résultats et au libre marché du choix de l'école par les parents), l'évaluation externe est accompagnée d'enjeux importants, sous la forme d'incitants. Les épreuves externes servent de reddition de compte externe et interne. On différencie généralement deux types d'épreuves externes selon les enjeux (*stakes*) qu'elles comportent pour les différents acteurs de la scolarité (Amrein & Berliner, 2002 ; Haney & Madaus, 1986 ; McDonnell, 1994) :
 - *High-stakes tests*. Les résultats de ces épreuves externes entraînent des enjeux importants pour les élèves, pour les enseignants, pour les établissements qui se voient octroyer des sanctions ou des récompenses¹⁶ ;
 - *Low-stakes tests*. Les résultats de ces épreuves externes ne débouchent directement sur aucune récompense ni sanction.

Pour McDonnell (1994), les *low-stakes tests* se localisent plutôt sur un « pôle informatif » et les *high-stakes tests* sur le plutôt sur un « pôle régulation ». Néanmoins, Monseur et Demeuse (2005) mettent en doute ce classement des épreuves sur ces deux pôles :

Cette classification peut néanmoins être questionnée : la pression induite par la publication de certains résultats, bien qu'à titre informatif, peut conduire à des pressions très fortes de la part du « marché scolaire » alors que des épreuves destinées à réguler le système peuvent avoir peu d'impact sur les élèves ou les écoles elles-mêmes, mais porter sur la définition du curriculum. Néanmoins, cette distinction reste utile lorsqu'on analyse les caractéristiques des épreuves en termes de pression exercées sur les différents acteurs (p.9).

Cette terminologie en termes d'enjeux (*stakes*) est donc spécifique à des pays qui utilisent ouvertement les tests standardisés comme moyen de pression dans un modèle d'*accountability* dure. Ces dispositifs sont créés dans ce but et prévoient des incitants (récompenses et sanctions). Il serait alors, selon nous, erroné d'utiliser cette terminologie pour des épreuves externes ancrées dans un contexte éducatif différent qui ne comprennent pas de telles politiques.

- Dans les contextes où les politiques d'*accountability* sont dites « douces », l'évaluation standardisée sert surtout au contrôle et à la « reddition de comptes » des entités locales décentralisées, toutefois de manière interne uniquement. En effet, dans ces contextes, les résultats des épreuves externes sont rarement publiés publiquement.

¹⁶ Les divers incitants sont décrits au point 3.4.4 (p.21-22).

L'analyse fine des dispositifs réellement mis en œuvre dans le cadre européen montre que ce n'est pas le principe de responsabilisation en direction de l'utilisateur parents ou citoyen qui prime mais bien le contrôle administratif interne (Mons & Crahay, 2011, p.94).

Les résultats des épreuves externes, même sans enjeux forts, entraînent également des conséquences sur les acteurs de terrain. Pour Huguenin & Soguel (2013), elles donnent au système scolaire une certaine transparence qui permet :

- Aux élèves de se situer, notamment en se comparant à leurs pairs ;
- Aux parents de situer la performance de leur enfant et d'identifier les acteurs qui sont responsables de leurs résultats ;
- Aux enseignants d'identifier le niveau d'exigence demandé et d'adapter leurs pratiques ;
- Aux directions d'école d'évaluer les enseignants et l'établissement ;
- Aux administrations d'identifier les établissements (et les enseignants) performants ou non.

5.3. Les fonctions pédagogiques de l'évaluation externe

La fonction originelle de l'évaluation externe des acquis des élèves est pédagogique. En tant qu'alternative ou de complément à l'évaluation interne, elle est un moyen de « développement » ou de « contrôle » de l'apprentissage des élèves (Ntamakiliro & Longchamp, 2008). L'évaluation externe peut donc avoir des fonctions semblables à celles de l'évaluation en classe (figure 3).

Évaluation des apprentissages

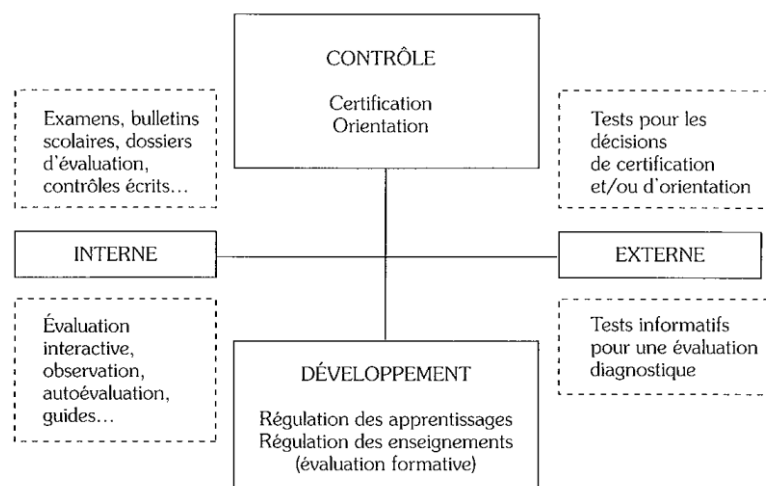


FIGURE 3 – Contextes de l'évaluation des apprentissages (Mottier Lopez, 2009, p.14)

5.3.1. Des fonctions pédagogiques semblables à celles de l'évaluation interne

L'évaluation externe tient plusieurs fonctions pédagogiques qui s'apparentent à celles de l'évaluation menée par l'enseignant dans sa classe (interne). Ces fonctions apparaissent sous de nombreux termes dans la littérature : pronostique, diagnostique, sommative, formative, certificative, ... Nous présentons ici leurs différences.

La typologie des fonctions de l'évaluation élaborée par Bloom, Madaus, & Hastings (1981) fait généralement office de référence dans le champ de l'évaluation scolaire (Allal, 1991 ; De Landsheere, 1979 ; Hadji, 1989 ; Scallon, 1988). Trois fonctions principales de l'évaluation sont distinguées, déterminées notamment par les intentions de l'évaluateur et du moment où elles interviennent dans le processus d'apprentissage :

- *Évaluation diagnostique*. Elle permet avant ou en début d'apprentissage d'identifier les forces et les faiblesses (mais aussi les prérequis) d'un élève dans le but d'adapter la séquence d'enseignement.
- *Évaluation formative*. Elle permet pendant l'apprentissage d'identifier les forces et les faiblesses d'un élève dans le but d'y remédier.
- *Évaluation sommative*. Elle permet après l'apprentissage de faire un bilan qui mesure le degré d'acquisition des objectifs par l'élève.

Toutefois, dans la littérature, viennent s'ajouter deux autres fonctions :

- *Évaluation certificative*. Elle permet à la fin de l'apprentissage de valider les acquis de l'élève dans un but d'accession à un niveau d'étude supérieur ou d'obtention d'un certificat. Elle peut aussi permettre d'orienter les élèves dans une filière (Roegiers, 2004). Pour cet auteur, la fonction "certificative" n'est pas synonyme de "sommative", cette dernière désignant le type de recueil d'information (qui permet de faire la somme des acquis des élèves de plusieurs apprentissages, correspondant à une somme d'objectifs). Toutefois, pour Scallon (1991) « la pratique de l'évaluation exclusivement terminale a cédé sa place à une pratique d'évaluation sommative dite continue, où les résultats partiels obtenus au long du trajet sont accumulés arithmétiquement (voire additionnés) pour constituer un bilan "sommatif" » (p.71).
- *Évaluation pronostique*. Elle permet de prédire la capacité d'un élève à réussir un apprentissage, voire un cycle d'études. Ce type d'évaluation sert généralement à sélectionner les élèves dans des voies d'études ou comme test d'entrée (Hadji, 1989).

Allal (1991) propose une représentation graphique des fonctions de l'évaluation scolaire et des décisions qui en résultent (figure 4) :

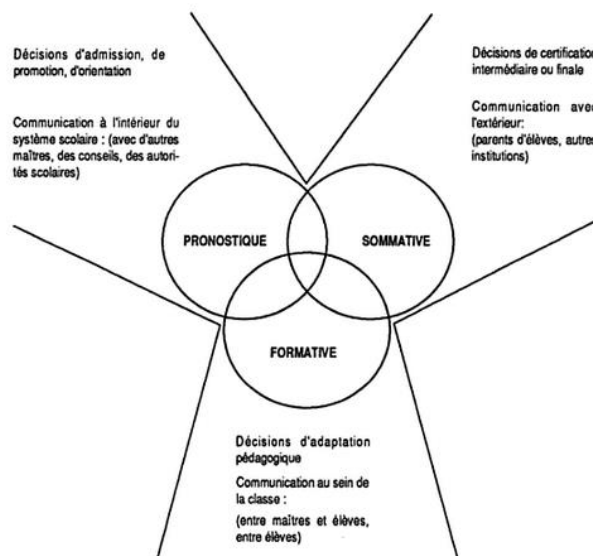


FIGURE 4 – Les fonctions de l'évaluation scolaire (Allal, 1991, p.12)

Roegiers (2004) distingue encore des différences entre évaluations selon leur interprétation qui peut être de deux natures :

- *Interprétation normative.* L'évaluateur détermine la performance de l'élève par rapport à celle des autres élèves.
- *Interprétation critériée.* L'évaluateur détermine la performance de l'élève par rapport à l'atteinte d'un ou de plusieurs objectifs à l'aide de critères.

5.3.2. Les fonctions pédagogiques de l'évaluation externe : catégorisations dans la littérature

A l'image de l'évaluation interne, l'évaluation externe des acquis des élèves peut être catégorisée selon ses fonctions pédagogiques. Ce sont les intentions de l'évaluateur et le moment où elles interviennent dans la période d'apprentissage qui déterminent ce classement. Plusieurs typologies sont proposées dans la littérature :

- Le rapport Eurydice (2009) différencie deux fonctions pédagogiques des épreuves standardisées nationales ou régionales :
 - o Évaluation externe « sommative » (évaluation de l'apprentissage). Ce type d'évaluation externe a pour fonction de faire le bilan des apprentissages des élèves à la fin d'un niveau de la scolarité. Ces tests présentent des enjeux importants pour la carrière scolaire de l'élève (certification, promotion, orientation). Ces tests ont généralement un caractère obligatoire.

- Évaluation externe « formative » (évaluation pour l'apprentissage). Ce type d'évaluation externe a pour fonction d'aider l'élève et/ou l'enseignant lors du processus d'apprentissage. Il permet d'identifier les besoins spécifiques des élèves et/ou d'adapter un enseignement. Ces tests présentent généralement des enjeux limités pour l'élève à titre individuel, mais « constituent des instruments importants pour améliorer les performances et peuvent conduire à des progrès importants en termes d'acquis pédagogiques » (p.24). Ces tests sont obligatoires ou facultatifs.
- Newton (2007) cite trois fonctions qui sont données à l'évaluation externe en Angleterre. Outre sa fonction politique (*evaluative uses : assessment for accountability*), il distingue trois fonction pédagogiques :
 - *Formative uses (assessment for learning)*
 - *Summative uses (assessment of learning)*
 - *Diagnostic uses (assessment for special intervention)*
- En Belgique francophone, les épreuves externes sont réparties en deux catégories par les autorités (Lafontaine, 2012) :
 - Épreuves externes « certificatives ». Elles amènent les élèves à l'obtention d'un certificat.
 - Épreuves externes « non certificatives ». Elles sont diagnostiques et visent à informer les enseignants à propos des acquis de leurs élèves.
- Soussi & Nidegger (2012), retiennent deux fonctions pédagogiques sans toutefois les définir :
 - Évaluation externe certificative
 - Évaluation externe formative-diagnostique
- Le paysage éducatif suisse et plus particulièrement romand compte de nombreuses évaluations externes, communément appelées épreuves cantonales. Chaque canton gère en effet ses propres dispositifs. Un classement en a été fait par le Consortium romand¹⁷ qui a eu pour mission d' « harmoniser les définitions des épreuves externes, afin de se repérer dans les divers types d'évaluations cantonales, et de s'accorder sur l'emploi d'une terminologie » (Marc & Wirthner, 2012, p. 11). Selon leur fonction, les épreuves externes ont été séparées en deux catégories (figure 5) :

¹⁷ Il s'agit d'un groupe de travail qui regroupe les responsables de l'évaluation des cantons de Suisse romande.

Évaluation à visée sommative: ce type d'évaluation regroupe ici autant des évaluations qui seront destinées à orienter les élèves que des évaluations ayant un impact dans la promotion des élèves.

« Une évaluation à visée sommative est une évaluation critériée qui permet de dresser un inventaire des compétences et des connaissances acquises par l'élève au terme d'une ou plusieurs séquences d'enseignement, correspondant à un ensemble d'objectifs d'apprentissage. »

Évaluation à visée diagnostique: ce type d'évaluation vise avant tout à informer les enseignants, les élèves et leurs parents sur l'état d'avancement des apprentissages dans le but de pallier tout manque éventuel. Ce type d'évaluation a été défini comme suit :

« Une évaluation à visée diagnostique a pour but de fournir aux enseignants des informations sur le degré d'atteinte de certains objectifs du plan d'études en complément des procédures d'évaluation qu'ils mènent par ailleurs, afin d'identifier les difficultés d'atteinte de ces objectifs et de conduire aux remédiations nécessaires. »

FIGURE 5 – Typologie des épreuves cantonales externes du Consortium romand (Marc & Wirthner, 2012, p.11)

- Pour Tessaro & Ntamakiliro (2010), classer les évaluations externes (notamment celles de Suisse romande) en termes de fonctions est une « question réductrice » car elles ne sont plus seulement ciblées sur l'évaluation des apprentissages des élèves mais aussi, et surtout, destinées à estimer le niveau des classes, des régions, des pays. Ils proposent plutôt de les distinguer en termes d'enjeux et différencient deux catégories :
 - o Évaluation externe comportant des « enjeux faibles » (*low-stakes tests*) pour les élèves, « subsidiairement » pour les enseignants et les autorités scolaires.
 - o Évaluation externe comportant des « enjeux élevés » (*high-stakes tests*) pour les élèves, « subsidiairement » pour les enseignants et les autorités scolaires.

Soussi & Nidegger (2012) remettent en question ce classement car il est inspiré du contexte politique américain bien particulier. Le terme de *stakes tests* est spécifique à des contextes d'*accountability* dur.

Précisons toutefois que la notion d'enjeux élevés doit être relativisée étant donné que les répercussions ne sont pas directes (les résultats des épreuves participent à la certification sans être déterminantes) ni du même type qu'en Amérique du Nord (p.3).

5.3.3. Les fonctions pédagogiques de l'évaluation externe : notre catégorisation

En nous référant aux différentes catégories proposées dans le point précédent, nous proposons notre propre typologie de l'évaluation externe des acquis des élèves :

- **Evaluation externe POUR l'apprentissage (développement de l'apprentissage)**
 - À visée diagnostique. L'épreuve externe sert à situer les apprentissages des élèves au début de la période d'apprentissage.
 - À visée formative. L'épreuve externe sert à situer et à réguler les apprentissages des élèves (et/ou l'enseignement) au cours de la période d'apprentissage.
- **Evaluation externe DES apprentissages (contrôle de l'apprentissage)**
 - À visée pronostique. L'épreuve externe sert à prédire (avant la période d'apprentissage) la capacité des élèves à réussir un apprentissage, voire un cycle d'études. Ce type d'évaluation sert notamment à sélectionner les élèves dans des voies d'études.
 - À visée sommative. L'épreuve externe sert à faire un bilan qui mesure le degré d'acquisition des objectifs par les élèves après une période d'apprentissage. Elle peut être un des éléments pris en compte dans la certification et l'orientation des élèves.
 - À visée certificative. L'épreuve externe sert à valider les apprentissages des élèves après une période d'apprentissage. Elle permet à elle seule l'accession à un niveau d'étude supérieur ou à l'obtention d'un certificat.

5.4. Les tensions entre les différentes fonctions de l'évaluation externe

Nous avons démontré que l'évaluation externe des acquis des élèves poursuit des fonctions politiques et pédagogiques. Il arrive souvent, notamment pour minimiser les coûts, qu'elles remplissent plusieurs fonctions simultanément, et qui sont parfois « contradictoires ». Peuvent alors apparaître des « tensions »¹⁸, voire des conflits (Mons & Pons, 2006).

¹⁸ Cette problématique est discutée par les contributions à l'ouvrage dirigé par Mottier Lopez & Crahay (2011) « Evaluations en tension : Entre la régulation des apprentissages et le pilotage des systèmes ».

5.4.1. Des tensions entre fonctions pédagogiques et politiques

Bien souvent dans les pays de l'OCDE, une même épreuve externe tient simultanément plusieurs fonctions qui touchent des niveaux différents : un rôle pédagogique (micro) et un rôle politique (macro). Par exemple, des épreuves diagnostiques servent également à dresser le bilan des acquis des élèves au niveau du système ou des épreuves conçues initialement pour mesurer les acquis des élèves servent parfois à responsabiliser les établissements ou les enseignants (Eurydice, 2009 ; Mons & Crahay, 2011). Pourtant, pour être valide, une évaluation devrait ne remplir qu'une unique fonction (Eurydice, 2009 ; Mons & Pons, 2006).

A couvrir trop de finalité à la fois, les épreuves standardisées risquent d'apporter des réponses fausses à des questions qui ne sont pas rigoureusement définies. En particulier, il paraît important de distinguer le *testing* pour la redevabilité et les épreuves servant dans le cadre de la régulation des apprentissages (Mons & Pons, 2006, p.129).

En effet, les résultats d'évaluations destinées à mesurer les acquis des élèves ne sont pas assez complets pour piloter un système entier. Plusieurs raisons sont avancées : 1) ces épreuves se limitent bien souvent à un nombre limité de disciplines dites "prioritaires" (mathématiques, langues, sciences naturelles) ; 2) les données sur les élèves sont insuffisantes, voire inexistantes ; 3) les données sur les établissements sont insuffisantes, voire inexistantes (Mons & Crahay, 2011). A ces raisons, nous ajoutons également que le traitement méthodologique et statistique des résultats des épreuves ne se fait pas toujours de manière scientifique. Finalement, les auteurs, s'appuyant sur le rapport Eurydice (2009), soulignent que les résultats se centrent souvent sur la performance des établissements (ou des classes) mais rarement sur celle des politiques. De plus, ces résultats sont collectés sur toute la population et non sur des échantillons représentatifs, ce qui pourrait permettre de réaliser de substantielles économies avec des résultats tout aussi valides (comme le font les études internationales). Toutefois, ce choix n'est pas anodin :

Il traduit un ciblage de l'évaluation des performances scolaires des élèves sur les établissements et les enseignants plutôt que sur l'évaluation des politiques nationales. Bien plus, on trouve peu d'opérations évaluatives qui ont pour finalité explicite de répondre à une question précise concernant une décision politique prise ou à prendre en regard de l'organisation de l'école (Mons & Crahay, 2011, p.88).

5.4.2. Des tensions entre fonctions pédagogiques

Certaines finalités pédagogiques de l'évaluation peuvent aussi être parfois en tension. Par exemple, des évaluations externes créées dans un but pronostique ou diagnostique sont parfois utilisées à des fins sommatives. Toutefois, un instrument d'évaluation devrait n'avoir qu'une unique finalité (Eurydice, 2009 ; Mons & Crahay, 2011).

De même, les deux niveaux d'évaluation (interne et externe) sont certes complémentaires, mais peuvent également se retrouver parfois « en tension ». En effet, ils peuvent être mis en confrontation, créant une hiérarchie où l'évaluation externe est considérée comme de meilleure qualité (même par les enseignants). Les pratiques évaluatives de l'enseignant sont alors dévaluées, considérée comme moins bonnes. Le « jugement professionnel » de l'enseignant est alors mis en doute. Pourtant, l'évaluation interne prend en compte bien plus de facteurs qualitatifs, tels que le comportement, l'engagement ou l'investissement de l'élève (Allal, 2009). Pour la plupart des chercheurs, l'évaluation interne et externe ne devrait pas être en opposition mais se compléter, tirant profits des avantages de chacune (Allal, 2009).

Au cours des 10 dernières années, plusieurs pays ou juridictions scolaires ont développé des épreuves externes, tant pour assurer une certaine régulation du système éducatif que pour permettre aux enseignants et aux établissements de mieux ajuster leurs interventions en fonction des exigences du curriculum. Un enjeu actuel est de renforcer la complémentarité entre les pratiques évaluatives des enseignants et les évaluations externes et d'améliorer chacune des formes d'évaluation en respectant leurs logiques propres (Mottier Lopez & Laveault, 2008, p. 21).

5.4.3. La nécessité de réduire ces tensions

En conclusion à ce point sur les tensions entre fonctions de l'évaluation externe, les constats présentés démontrent que chaque évaluation externe devrait idéalement suivre une seule finalité.

Les débats d'experts sur l'évaluation ont mis en lumière le fait que le recours à un seul et unique test à plusieurs fins pourrait se traduire par des contradictions quant aux exigences en termes d'information qu'implique chaque objectif. Dans ce cas, il est préconisé que les autorités éducatives établissent un ordre de priorité parmi les différents objectifs et ajustent la conception des tests en conséquence » (Eurydice, 2009, p.24).

Si cette condition n'est pas remplie, on court le risque que ces évaluations soient mal comprises par les décideurs mais aussi par les acteurs de terrain. Peuvent s'en suivre des effets néfastes et inattendus. Pour Mottier Lopez (2009), une épreuve externe devrait n'avoir qu'une seule cible, politique ou pédagogique, sans se préoccuper des autres niveaux de contexte. Toutefois, au niveau pratique, cette optique ne semble pas tout à fait possible ni pertinente.

Cette option est évidemment contraire à une approche systémique des enjeux de l'évaluation en éducation. De plus, elle ignore la réalité du terrain, les acteurs étant aujourd'hui confrontés aux différents types d'évaluation et à des modes de fonctionnement et d'organisation du travail : ce sont les enseignants par exemple qui sont responsables de la passation des épreuves externes (p.13).

Pour Crahay (2009), en conclusion à l'ouvrage « Evaluations en tensions », il n'est « pas impossible » de faire coexister les différentes fonctions de l'évaluation externe, étant toutes deux nécessaires pour l'intérêt général (au niveau macro) et local (pour le niveau micro). Toutefois, ces niveaux doivent

fonctionner en se complétant, formant une sorte de pilotage « bienveillant » de type plutôt informatif que certificatif. Il relève justement l'importance du niveau méso, intermédiaire entre la politique et le terrain. Laveault (2009b) propose d'ailleurs l'« évaluation intérimaire (*interim assessment*) » comme outil d'information des différents acteurs de tous niveaux. Il s'agit d'évaluations externes réalisées fréquemment (à plusieurs moments d'une année scolaire) qui font le bilan des acquisitions des élèves.

L'évaluation intérimaire se veut formative en permettant une régulation plus fréquente des systèmes éducatifs à tous les niveaux et en intégrant ce type de régulations à des régulations plus fines se produisant en salle de classe. Pour permettre de profiter de ce type de régulations plus fréquentes, il faut disposer d'épreuves qui fournissent des résultats valides à tous les niveaux du système éducatif. [...] Dans un tel contexte, le rôle des décideurs et des enseignants est davantage celui de consommateurs ou concepteurs éclairés de ces épreuves et d'utilisateurs prudents et compétents des résultats de ces épreuves (Laveault, 2009b, p. 12).

6. Effets de l'évaluation externe des acquis des élèves

Comme le montre notre développement dans les points précédents, l'évaluation externe des acquis des élèves a été instaurée dans les systèmes éducatifs prioritairement dans le but premier d'améliorer la qualité de ceux-ci (sous-entendu la performance des élèves). Toutefois, cet outil pédagogique et politique entraîne bien d'autres effets, attendus ou non, bénéfiques ou non, sur le système dans sa globalité mais également sur chacun des acteurs qui le compose à tous les niveaux (élèves, parents, enseignants, établissement).

Les différents acteurs scolaires – des pouvoirs publics jusqu'aux enseignants en passant par les acteurs intermédiaires (inspecteurs) – sont supposés utiliser les résultats de ces évaluations externes pour améliorer leurs performances à leur niveau (Dumay, Cattonar, Maroy, & Mangez, 2013, p. 136).

Ces effets sont généralement reconnus par la littérature. Toutefois, leur catégorisation ne fait pas forcément consensus : selon les auteurs, ils peuvent être classés en termes de gain vs. perte, effets positifs vs. effets négatifs, effets néfastes/pervers vs. effets bénéfiques. Cependant, ce qui est bénéfique (ou positif) pour certains n'est pas forcément bénéfique pour d'autres. L'objectif de la présente partie est de recenser tous les effets qui sont traités dans la littérature scientifique. Nous avons pris le parti de ne pas les classer en termes de valeur *a priori*, mais d'aborder tout d'abord les effets qui visent le système dans sa globalité, puis de les classer selon les acteurs visés. Nous détaillons ces effets et présentons des recherches qui les identifient.

6.1. Les effets de l'évaluation externe sur le système éducatif

Certains effets de l'évaluation externe des acquis des élèves touchent le système éducatif dans sa globalité. Elle doit servir à améliorer la performance du système éducatif, elle doit garantir une évaluation de qualité et faire diminuer les inégalités scolaires. Nous exposons ces trois points.

6.1.1. Amélioration de la performance du système éducatif

Le principal effet attendu de l'évaluation externe est d'améliorer la qualité d'un système éducatif. Elle doit amener à l'amélioration des performances des élèves (critère principalement évoqué). C'est surtout dans la littérature américaine que cet effet a été interrogé, ayant été le sujet de vifs débats politiques et scientifiques dans le cadre de la mise en place des politiques d'*accountability*. Ces études démontrent qu'aucun consensus n'est fait sur cette question, notamment à cause de biais méthodologiques (Mons, 2009). Nous détaillons certaines de ces études.

Les premières de ces études ont été menées dans le contexte des évaluations standardisées du Texas. Dans cet état américain, un dispositif de *testing*, nommé *Texas assesement of academic skills (TAAS)*,

a été mis sur pied pour « responsabiliser » les élèves et les écoles. Ce dispositif prévoit le classement des écoles dans des catégories en fonction des résultats de leurs élèves à un test standardisé (*high stakes test*). A ces catégories sont liées des récompenses ou des sanctions ayant des conséquences importantes sur la carrière des élèves (certification, redoublement), sur les enseignants (primes, renvoi) et sur les établissements (réputation, financement, fermeture) (Mons, 2009). Suite à l'implantation de ces politiques, une étude a révélé l'amélioration des performances des élèves, grâce à l'analyse des résultats de ceux-ci au TAAS (mais également en Caroline du Nord) (Grissmer & Flanagan, 1998). Cette recherche conclut notamment que ces gains de performance sont dus aux politiques mises en place dans ces deux Etats : la standardisation du curriculum, leur évaluation par un test standardisé et la mise en place d'incitants. Toutefois, ces résultats furent remis en doute par d'autres chercheurs qui dénoncèrent le « mythe texan », mettant en cause les méthodes d'analyse utilisées (Haney, 2000). En effet, en comparant les résultats du test texan et ceux du test national américain (*NAEP*) et en les analysant sur une longue période, on découvrit que les progrès avancés étaient en réalité minimes voire non significatifs (Fuller & Treisman, 2001). De plus, en analysant ces résultats sur le long terme, on s'aperçoit que les progrès aux tests locaux à forts enjeux existent bel et bien dans les premières années de mise en place mais disparaissent ensuite. Cette variabilité temporelle est expliquée par l'influence des pratiques des enseignants, notamment de techniques d'entraînement intensif pour la réussite du test (*teaching to the test*)¹⁹ (Nichols, 2007). En outre, les bons résultats s'expliquent en partie par l'exclusion des élèves faibles de l'échantillon (Haney, 2000).

Les dispositifs d'*accountability* s'étant généralisés dans tout le pays, les chercheurs ont pu comparer cet effet entre les Etats américains. Ainsi, Amrein & Berliner (2002) ont comparé les résultats de différents tests nationaux américains (*NAEP, ACT, SAT, AP*) selon les politiques d'*accountability* (*high-stakes testing*) mis en place par 18 Etats. Leurs résultats montrent que ces politiques n'ont au mieux aucun effet sur les performances des élèves, au pire les font baisser. En outre, leurs conclusions mettent en lumière des effets inattendus et néfastes de telles politiques de *testing* (*drop-out*, tricheries, démotivation des enseignants...). Ces résultats furent mis en doute par Rosenshine (2003) qui, après avoir analysé les mêmes données avec une méthode différente, avança des résultats montrant que les politiques de *testing*, surtout à forts enjeux, amélioraient les performances des élèves. Reprenant sa méthode, Amrein-Beardsley & Berliner (2003), purent toutefois démontrer que cet effet était nul si l'on prenait en compte les élèves exclus du test.

D'autres chercheurs affinèrent ces comparaisons en hiérarchisant les politiques d'*accountability* des Etats (classées de *low* à *high*). Encore une fois, les résultats de ces études ne trouvent pas de consensus. Ainsi, toutes menées sur une vingtaine d'états, certaines recherches montrent que les

¹⁹ Cet effet est l'objet du point 6.2.4 (p.66-68).

dispositifs de *testing* à forts enjeux produisent des gains de performance (Carnoy & Loeb, 2002; Hanushek & Raymond, 2004). D'autres montrent le contraire, pointant notamment l'inconsistance des résultats, les corrélations entre les enjeux et les performances n'étant pas régulières (parfois positives, parfois négatives) (Nichols, Glass, & Berliner, 2006).

Vu ces résultats contradictoires, une méta-analyse de ces recherches a été réalisée par Lee (2008). Ce chercheur américain a sélectionné 14 recherches (dont font partie celles citées auparavant) sur la base de critères méthodologiques : contexte de *high stakes testing* ; branches similaires (maths et lecture) ; exclusion des tests locaux ; échantillons représentatifs. Les résultats de cette méta-analyse (après contrôle des différents biais méthodologiques éventuels) ne montrent aucun effet positif des politiques de *high stakes tests* sur les performances des élèves, même en isolant certaines variables comme la branche ou les degrés scolaires.

Peu de recherches explorent cet effet hors du contexte américain. Toutefois, Bishop (1995), en se basant sur les données des études internationales *International Assessment of Educational Progress (IAEP)*, démontre que les systèmes éducatifs nationaux intégrant des épreuves externes montrent de meilleures performances en mathématiques (tout en tenant compte du niveau socio-économique). De même, Wößmann (2007), économiste allemand, démontre l'effet bénéfique des politiques d'*accountability* (et de compétition) au niveau international. Ses analyses multi-niveaux se basent sur la comparaison entre les résultats de quatre célèbres épreuves internationales (PISA, PIRLS, TIMSS et TIMSS-R) et les politiques éducatives des pays. Les résultats de ces comparaisons font ressortir les caractéristiques des systèmes où les élèves sont les plus performants. Parmi ceux-ci, apparaissent l'importance des évaluations externes associées à des mesures d'*accountability*. Ces effets positifs sont d'autant plus importants si ces mesures sont combinées avec une large autonomie des établissements et des enseignants.

A nouveau, ces résultats ne font pas l'unanimité. Ils sont en effet mis en doute par l'étude de Mons (2007). Se basant en partie sur les mêmes données (PISA, 2000), elle confirme tout d'abord les résultats de Wößmann (2007). Toutefois, en ajoutant comme variable de contrôle le PIB par habitant, elle démontre que les bons résultats s'expliquent surtout par le niveau socio-économique et non par l'usage d'évaluations standardisées comme avancé. L'étude de l'OCDE (2007) sur PISA 2006 parvient d'ailleurs aux mêmes conclusions : aucun lien significatif n'est trouvé entre les politiques de *testing* et une meilleure performance des élèves.

En conclusion, on admettra qu'à ce jour aucun consensus n'est atteint sur le fait que l'évaluation standardisée améliore ou non la performance (donc la qualité) des systèmes éducatifs. Ce jugement est partagé par Mons (2009) :

En conclusion, que l'on examine la littérature qui s'est concentrée sur les cas nationaux ou celle qui s'est appuyée sur les comparaisons internationales, les relations entre efficacité et *testing* apparaissent peu consistantes ; elles ne sont ni univoques, ni automatiques (p.117).

De plus, peu d'études s'intéressent à l'efficacité de ces politiques de testing, en mettant en balance leurs coûts et leurs résultats.

6.1.2. Evaluation garante de qualité

Dans la plupart des systèmes scolaires, c'est surtout l'évaluation de l'enseignant (interne) qui détermine la promotion ou non des élèves. Cependant, de nombreux biais bien connus peuvent entacher les pratiques évaluatives des enseignants et désavantager certains types d'élèves : effet Posthumus, culture de l'échec, etc. (Roegiers, 2004). Les décisions de promotion sont donc parfois injustes si elles se basent uniquement sur les évaluations internes, la promotion ne dépendant pas uniquement du niveau de performance de l'élève intrinsèquement, mais d'autres facteurs tels que la comparaison entre élèves par exemple. D'autres moyens sont donc nécessaires pour s'assurer l'objectivité et l'équité de l'évaluation. Dans la plupart des systèmes scolaires, il est attendu par les autorités que l'évaluation externe tienne ce rôle.

En effet, il est généralement accepté, même par les enseignants²⁰, que les épreuves externes permettent une évaluation de qualité des apprentissages des élèves. Ainsi, les épreuves externes ont souvent le rôle de certification (et d'orientation) des élèves dans les systèmes scolaires. Elles servent à certifier (voire orienter) les élèves avec précision (Eurydice, 2009). L'étude de Baeriswyl, Trautwein, Wandeler, & Lüdtke (2009), basée sur le processus d'orientation des élèves au niveau secondaire dans le canton de Fribourg (partie germanophone), montre que les épreuves externes standardisées ont un pouvoir pronostique très élevé (tout de même suivi de très près par l'évaluation interne de l'enseignant). Elles permettraient alors une orientation des élèves de qualité. Les chercheurs relèvent toutefois les risques d'effets néfastes lorsque l'importance donnée au test est trop importante. Ce constat est confirmé par Grisay (1984). Elle a mené en Belgique francophone une étude sur 53 écoles, comparant les résultats d'élèves à une épreuve externe de français et à une évaluation interne portant sur les mêmes objectifs du programme scolaire. Cette comparaison montre que la dispersion des notes de l'évaluation interne est plus importante que celle de l'évaluation externe. Toutefois, il a été observé que le classement des élèves reste sensiblement le même entre l'évaluation interne et externe.

²⁰ Les perceptions des enseignants quant à l'évaluation externe sont l'objet du point 8 (p.83-84).

6.1.3. Lutte contre les inégalités scolaires

Il est également attendu des épreuves externes qu'elles soient un moyen efficace pour lutter contre les inégalités. En effet, elles sont une mesure universelle, elles ne font pas de différence entre les élèves et évitent les systèmes élitaires favorisant les couches sociales les plus élevées. Outre le fait qu'elles soient un vecteur d'objectivité et d'impartialité, elles obligent également les enseignants à respecter les programmes curriculaires²¹ (le non-respect des programmes ayant été identifié comme source d'échec des élèves des minorités) (Crahay & Lafontaine, 2012). Plusieurs recherches suisses se sont penchées sur cette problématique. Ramirez (2011), comparant les résultats d'élèves du canton de Genève en classe et à une épreuve cantonale externe, démontre que les épreuves externes permettent d'éviter les biais évaluatifs dus à la relation élève-enseignant. Les épreuves cantonales permettraient ainsi de réduire les inégalités. Aussi Baeriswyl, Wandeler, Trautwein et Oswald (2006) démontrent, par l'analyse du système d'orientation des élèves du canton de Fribourg (partie germanophone), que l'usage d'une épreuve standardisée comme critère réduit l'influence de l'origine familiale de l'élève.

Toutefois, on peut parfois considérer à tort que les épreuves soit une mesure « objective et juste ». En effet, « Tout test, aussi bien construit soit-il, est passible d'erreur de mesure » (Crahay, 2009, p.237). Certaines études montrent justement que les épreuves externes connaissent également leur lot de biais évaluatifs. Petrucci (2010) démontre que les performances des élèves aux épreuves cantonales genevoises de lecture dépendent, outre de leurs caractéristiques personnelles, également de la classe dans laquelle ils évoluent (entre 10% et 18% de la variance, selon le degré scolaire). Cet effet-classe serait en grande partie expliqué par l'influence de l'enseignant sur les acquisitions de ses élèves. Ce constat est partagé par les études menées sur l'épreuve d'orientation du canton de Fribourg (partie germanophone). Celles-ci montrent des effets de classe, mais aussi de l'école et de l'enseignant (Baeriswyl, Wandeler, & Trautwein, 2011; Baeriswyl et al., 2006). Aussi, se référant aux épreuves cantonales du canton de Vaud, Ntamakiliro & Longchamp (2008) observent des disparités de résultats entre établissements. Elles seraient une « plus-value » pour les élèves des établissements performants mais pénaliseraient les moins bien lotis. Enfin, ces épreuves comportent aussi des biais culturels. Elles sont ancrées dans un contexte culturel propre à celui de leur concepteur. Les items, les textes de lecture, les réponses attendues et les contenus choisis ne sont pas culturellement neutres, ni équitables. (Madaus, 1995 ; Monseur et Demeuse, 2005).

Dans certains contextes, les politiques de *testing* sont accompagnées d'actions visant à réduire les inégalités entre élèves. Toutefois, à nouveau, aucun consensus scientifique et empiriquement prouvé n'est à ce jour atteint, ni au niveau national ni international (Mons, 2013). Dans le contexte américain des *high stakes tests*, des mesures imposés par des lois fédérales (citées auparavant) poussent les Etats

²¹ Cet effet est l'objet du point 6.2.1 (p.54-55).

à réduire les écarts entre les élèves d'origines diverses. Certaines études tendent à prouver que les politiques des Etats, surtout lorsqu'elles sont accompagnées de forts enjeux (conséquences pour les élèves et/ou les enseignants), permettent de réduire ces inégalités dues « aux provenances ethniques » (Carnoy & Loeb, 2002; Grissmer & Flanagan, 1998; Hong & Youngs, 2008). Toutefois, les résultats de ces études ont été contestés, notamment à cause de leur méthodologie. D'autres chercheurs, travaillant pourtant sur les mêmes données, ne trouvent aucun effet positif (Darling-Hammond, 1994 ; Fuller & Treisman, 2001; Haney, 2000; Hanushek & Raymond, 2004; Nichols et al., 2006). La méta-analyse de Lee (2008) ne montre également aucun lien entre *testing* et réduction des inégalités.

Sur le plan international, le rapport de l'OCDE (2007) sur les données de PISA 2006 conclut à une absence d'un tel effet. D'autres études trouvent, au contraire, des corrélations positives entre épreuves nationales externes et réduction des inégalités, dans l'enseignement secondaire surtout (Harris & Herrington, 2006; Mons, 2007a). Toutefois, ces épreuves ne sont pas forcément intégrées dans une politique plus large d'*accountability*.

L'existence d'épreuves externes nationales pourrait agir ainsi moins parce qu'elles exercent des pressions psychologiques sur les élèves et les enseignants mais parce qu'elles homogénéisent les exigences académiques qui sont imposées aux enseignants et limitent ainsi les dérives curriculaires dans les établissements défavorisés (Mons, 2013, p. 44).

6.2. Les effets de l'évaluation externe sur les enseignants

Les effets des épreuves externes touchent également les enseignants, leurs pratiques et leurs attitudes. Ils sont attendus ou non par les autorités. Nous décrivons les effets identifiés par la littérature et les recherches empiriques.

6.2.1. Alignement des contenus enseignés

Un des effets attendus des évaluations standardisées par les décideurs est de clarifier les contenus qui doivent être enseignés mais aussi de préciser leur niveau d'atteinte. Elles précisent aux enseignants ce qui est attendu des élèves à certains moments de la scolarité, permettant alors l'« alignement » des contenus enseignés (Looney, 2009).

Dans le meilleur des scénarios, ces épreuves constituent une référence concrète sur les objectifs des programmes. Elles permettent d'améliorer l'alignement de l'enseignement et des évaluations en classe sur les objectifs des programmes d'études et d'accroître ainsi leur validité de contenu [...] Dans la mesure où un alignement sur les programmes d'études permet aux enseignants de mieux établir la progression des élèves dans le programme de formation, d'éviter les chevauchements et de préparer adéquatement les élèves aux niveaux suivants, les épreuves externes peuvent contribuer indirectement à améliorer les résultats d'apprentissage (Laveault, 2009b, p. 11 - 12).

Pourtant, d'autres moyens plus spécifiques sont généralement prévus pour formuler et communiquer ces contenus : les programmes, plans d'études, standards de formation, moyens d'enseignement. Toutefois, ces outils seraient moins efficaces que les épreuves externes.

Quoiqu'il en soit, les programmes n'ont pas force de loi chez les praticiens de l'éducation et il semble vain d'espérer changer les pratiques par la simple promulgation de listes d'objectifs ou de socles de compétences ou même par la publication d'un curriculum structuré (Crahay & Lafontaine, 2012, p. 465).

De plus, les plans d'études ne définiraient pas suffisamment précisément le niveau attendu des élèves (Tessaro & Ntamakiliro, 2010a). Selon Mons & Crahay (2011) les épreuves externes seraient même plus efficaces que les standards de formation, étant davantage utilisés par les enseignants. Demailly (2005) insiste également sur le fait que les standards de formation peuvent indiquer le niveau attendu, mais les programmes scolaires doivent indiquer comment enseigner et apprendre. L'évaluation standardisée serait alors le meilleur moyen pour communiquer « l'attendu » aux enseignants et pour le contrôler (Carette, 2007).

[...] un outil d'évaluation qui traduit concrètement sous la forme d'une épreuve les attentes énoncées dans les référentiels peut se révéler utile. En effet, en se penchant sur des épreuves d'évaluation, les enseignants sont à même d'analyser l'interprétation de « l'attendu » réalisé par les concepteurs des épreuves. De ce fait, diffuser des épreuves d'évaluation, c'est diffuser une interprétation de « l'attendu » qui peut aider les enseignants dans leur propre interprétation. (Carette, 2007, p. 55).

Les épreuves externes permettent aux enseignants d'adapter leur enseignement par rapport à ce que les concepteurs des programmes attendent d'eux. Avec de telles épreuves, les enseignants peuvent observer les objectifs à atteindre et les compétences à maîtriser par les élèves, puis de les vérifier en comparant les résultats de leurs élèves avec ceux des autres classes. Ils peuvent ainsi déterminer les domaines où ils doivent s'améliorer (autant l'enseignant que les élèves) (Lafontaine et al., 2009; Monseur & Demeuse, 2005). En Angleterre, les politiques de *testing* permettent aux enseignants d'avoir une vision claire de ce qui est attendu des élèves à chaque niveau scolaire, il s'agit des *key stages* (Mons, 2009). Dans le canton de Genève, des enseignants affirment que les épreuves cantonales leur donnent des « repères précis et objectifs (Soussi, Ducrey, Ferrez, Nidegger, & Viry, 2006)²². Il en va de même dans le canton de Vaud (Longchamp, 2010)²³.

²² Un échantillon de 1640 enseignants genevois (du primaire, du CO et du post-obligatoire) a répondu à questionnaire concernant leurs pratiques d'évaluation, constitué d'échelles de Likert de 1 (pas du tout d'accord) à 6 (tout à fait d'accord).

²³ 20 enseignants de mathématiques du canton de Vaud ont été interrogés par entretien sur leur point de vue concernant l'utilisation des épreuves cantonales de mathématique de 6^{ème} année.

6.2.2. *Modification du curriculum enseigné*

Les résultats des évaluations externes informent les autorités sur le degré d'atteinte des objectifs fixés par les référentiels locaux ou nationaux. Ils permettent donc d'identifier les contenus qui sont à développer prioritairement dans les apprentissages des élèves (et/ou dans leur enseignement) (Mons & Crahay, 2011). Par le biais d'épreuves externes, les autorités peuvent ainsi imposer une plus grande attention sur certains contenus. C'est alors l'évaluation qui définit ce qui doit être enseigné et appris, davantage que les référentiels. Cet effet est appelé *washback* ou *backwash effect* (Alderson & Wall, 1992 ; Pan, 2009 ; Mizutani, 2009 ; Spratt, 2005). Ces termes sont surtout utilisés dans le champ de la linguistique, peu dans celui de l'enseignement en général. Il désigne le fait que l'évaluation externe standardisée (les tests standardisés d'anglais notamment) peut influencer de manière intentionnelle ou non ce qui est enseigné et appris (les contenus) et comment (les pratiques)²⁴. Monseur et Demeuse (2005) traduisent ce terme en français pas celui de « reflux ».

(L'effet de reflux) réside dans l'influence que les décideurs peuvent exercer sur le curriculum tel qu'il est enseigné. En associant à une épreuve externe des sanctions ou des récompenses, ils peuvent déterminer les finalités poursuivies par les professeurs et, dans une moindre mesure, les pratiques utilisées pour atteindre celles-ci (Monseur & Demeuse, 2005, p. 501).

Ces sanctions peuvent être de deux ordres : le premier est de pouvoir mettre une certaine pression sur les enseignants pour que leurs élèves atteignent les objectifs souhaités avec de bons résultats. Le deuxième consiste à inciter les enseignants à consacrer beaucoup de temps à préparer leurs élèves à réussir ces épreuves, surtout lorsque les résultats de l'évaluation externe comportent des conséquences pour l'élève (dans le cas notamment d'une épreuve certificative, de promotion ou d'orientation).

L'« effet de reflux » des évaluations externes peut de ce fait aider les autorités à introduire des changements dans les contenus et les faire adopter par les enseignants. Un tel phénomène a été observé dans de nombreux contextes. En France par exemple, les épreuves diagnostiques nationales, même sans enjeux, ont une « force normative ». Elles ont eu par exemple pour effet de renforcer l'attention des enseignants sur la géométrie (Mons & Pons, 2006). Aussi, dans l'état de New York, les autorités connaissaient peu de succès dans l'adoption d'une nouvelle manière de travailler les langues (en mettant l'accent sur la communication) jusqu'au jour où cette manière de travailler fut prise en compte dans le *State Regents Examination* (épreuve externe d'Etat). Les enseignants adaptèrent alors leur enseignement en prévision de ce test (Haney & Madaus, 1986). Dans le champ de la linguistique (*washback effect*), certaines études ont démontré que les tests standardisés peuvent avoir un effet (d'importance variable) sur le curriculum enseigné en classe comme, par exemple, la centration sur des objectifs de communication (Cheng, 1999; Shohamy, 2001).

²⁴ Les effets de l'évaluation externe sur les pratiques des enseignants sont l'objet du point 6.2 (p.54-72).

Les épreuves externes peuvent donc bien servir de levier lors de l'introduction d'un nouveau curriculum ou d'une réforme pédagogique, qu'il s'agisse d'un effet recherché ou incident (Monseur & Demeuse, 2005, p. 502).

Néanmoins, cet effet de reflux sur les contenus peut aussi être négatif. Premièrement, l'évaluation externe peut inciter les enseignants à traiter prioritairement les contenus qui sont évalués dans les épreuves externes. Cet effet est communément appelé rétrécissement du curriculum²⁵. Deuxièmement, par son caractère standard, elle entraîne le risque que les apprentissages et les contenus enseignés soient décontextualisés. En effet, il se peut que les évaluations externes comportent des contenus « aseptisés, détachés de tout contexte particulier » (Monseur & Demeuse, 2005, p.505). Cet effet est surtout valable pour les épreuves standardisées à un pays ou à de grandes régions. Il est plus faible lorsque l'épreuve touche un territoire plus restreint.

Ce qui est important pour une personne qui vit dans une région particulière ou qui fait partie d'une communauté donnée peut revêtir une toute autre importance pour une autre qui vit ailleurs. [...] Pour être efficaces, les enseignants doivent tenir compte des élèves qu'ils ont en face d'eux et non d'élèves tels que le programme idéal les aurait imaginés » (Darling-Hammond, 1994, cité par Monseur & Demeuse, 2005, p.505).

6.2.3. Régulation des pratiques des enseignants

Les autorités peuvent attendre que l'évaluation externe soit un outil de régulation²⁶ des pratiques des enseignants. Dans un système décentralisé où écoles et enseignants ont gagné en autonomie, l'évaluation standardisée permet de contrôler et de diriger les pratiques du terrain. L'évaluation externe peut être un « levier didactique » utile pour faire évoluer les pratiques des enseignants en classe vers les « bonnes pratiques » (Carette & Dupriez, 2009 ; Linn, 2000 ; Lafontaine, 2000 ; Laveault, 2009 ; Maroy, 2004 ; Mons & Crahay, 2011 ; Ntamakiliro & Tessaro, 2010).

Alors que la thèse d'un découplage entre les pratiques d'enseignement et les structures organisationnelles des écoles et systèmes éducatifs constitue une représentation de plus en plus répandue du fonctionnement des organisations scolaires, plusieurs études récentes avancent que les politiques d'accountability pourraient réussir à réaliser un de leurs objectifs affichés qui est justement de renforcer le couplage entre les pratiques des enseignants, les mécanismes formels d'évaluation externe et les standards de résultats visés (Dumay et al., 2013, p. 126).

Pourtant, à nouveau, les pratiques pédagogiques sont historiquement institutionnalisées par le biais des programmes, des plans d'études, des moyens d'enseignement, et surtout par la formation initiale et continue des enseignants. Mais ces outils seraient moins efficaces que l'évaluation standardisée pour communiquer et contrôler « l'attendu » (Carette, 2007), ici non sous forme de contenus mais sous

²⁵ Cet effet est l'objet du point 6.2.5 (p.68).

²⁶ L'évaluation externe comme outil de régulation est l'objet du point 5.2.2. (p.36-37).

forme de « bonnes » pratiques. L'évaluation standardisée étant considérée comme une mesure fiable, même par les enseignants, ceux-ci la prennent comme modèle dans le but d'amélioration de leurs pratiques d'enseignement et d'évaluation, dans la recherche d'une certaine validité de leurs pratiques (Laveault, 2009b). De plus, cet effet serait plus important si l'évaluation externe est incluse dans un système d'enjeux élevés (Popham, 1987).

Toutefois, nous souhaitons interroger le fait que l'évaluation externe permette une régulation efficace des pratiques des enseignants. Selon nous, l'interprétation solitaire des résultats d'épreuves externes par les enseignants ne suffit pas à la bonne compréhension de leurs résultats et à la mise en pratique de « l'attendu ». Une régulation réussie nécessite un réel feedback et des actions de remédiation²⁷ (Allal, 2007; Laveault, 2007; Sadler, 1989). Des outils complémentaires (analyses détaillées des résultats, formation continue, échanges en équipe, etc.) sont nécessaires pour accompagner l'interprétation de ses résultats et sa mise en pratique. Justement, certaines recherches démontrent que les changements attendus ne s'opèrent pas dans les classes. Plusieurs recherches portent sur l'impact des épreuves externes sur les pratiques des enseignants dans différents contextes, sans forcément trouver un consensus. Nous en faisons ici la recension :

- *En Belgique francophone.* Quelques 2000 enseignants belges (du primaire et du secondaire) ont été interrogés par questionnaire sur leurs usages de l'évaluation externe. Une (faible) majorité des enseignants primaire (55%) et une minorité pour le secondaire (25%) sont d'accord avec l'affirmation selon laquelle « L'évaluation externe a des retombées positives sur leur quotidien en classe ». Toutefois, l'auteur de la recherche s'interroge sur les usages effectifs des évaluations externes en classe (Lafontaine et al., 2009). Une étude menée par Ouzzine (2009, cité par Carette & Dupriez, 2009), interrogeant 36 enseignants belges (du primaire et du secondaire), montre qu'une partie de ceux-ci adaptent leur enseignement sur le modèle des épreuves externes. Au contraire, pour Carette & Dupriez (2009) en CFB également, bien que les évaluations externes ont permis de donner de nombreuses informations sur le système éducatif, leur rôle de régulation des pratiques n'a eu que peu de succès (en tout cas aucun élément ne le prouve ou son contraire).
- *En France.* Une enquête par questionnaire de la DEP (1994) montre que 67% des enseignants de CE2 utilisent l'évaluation externe nationale pour modifier leur enseignement. Étonnamment, dans la même enquête, trois quarts d'entre eux affirment ne pas voir l'utilité de ces épreuves pour leurs pratiques (Pons, 2011). L'effet de régulation des pratiques est mis en doute : les enseignants auraient un message de complaisance dans ce type d'enquête et dans

²⁷ Voir à ce sujet les apports théoriques au point 5.2.2. (p.36-37).

les faits auraient du mal et peu de volonté à utiliser les évaluations externes nationales dans leurs pratiques (Normand & Derouet, 2011).

- *En Suisse*. Quelques études se sont interrogées sur les effets des épreuves cantonales sur les pratiques des enseignants. Elles montrent que ces dispositifs externes peuvent avoir une influence sur les pratiques des enseignants. Des entretiens menés avec des enseignants vaudois montrent toutefois qu'ils sont plutôt partagés (Longchamp, 2010). Le rapport de recherche de Soussi et al. (2006) montre que les enseignants genevois sont plutôt partagés quant aux rôles positifs des épreuves cantonales. Les résultats du questionnaire montrent qu'elles peuvent, dans une certaine mesure, être acceptées par les enseignants (sans différence entre les degrés) comme génératrices de nouvelles idées et peuvent réguler leurs pratiques. Ces constats sont d'ailleurs partagés par l'étude de Ntamakiliro & Tessaro (2002) menée par entretien auprès d'enseignants genevois du degré primaire.
- *Aux USA*. Des chercheurs se sont intéressés à l'impact des *high stakes tests* sur les pratiques des enseignants. Dans ce contexte particulier, l'influence des tests semble très importante. Toutefois, la plupart des constats sont alarmants. Les épreuves conduisent plutôt les enseignants à adopter des méthodes d'évaluation et d'enseignement proches de celles des tests dans le but premier de préparer spécifiquement les élèves à leur réussite (*teaching to the test*) et à traiter en priorité les contenus évalués (rétrécissement du curriculum) (Au, 2007; Darling-Hammond & Wise, 1985; McMillan, 2005; Shepard, 2003). L'étude de Au (2007)²⁸ recense une minorité d'études qui montrent des effets positifs sur les pratiques de classe tels que le développement des contenus curriculaires (14 recherches sur 41), d'une pédagogie centrée sur l'apprenant et coopérative (10/34) et de l'interdisciplinarité (6/38).
- Cette problématique a été explorée dans la littérature sur le *washback effect* en utilisant des méthodes diverses (questionnaires, observations, entretiens) et dans divers contextes (USA, Chine, Sri Lanka, Hong Kong...). Certaines recherches montrent que les enseignants « alignent » bel et bien leurs pratiques sur celles présentées par les épreuves externes, mais que ces changements sont superficiels (par exemple : adoption uniquement du « format » des tests pour l'évaluation interne) (Shohamy, 2001; Smith, 1991). D'autres recherches montrent que ces intentions de régulation échouent, voire créent d'autres effets indésirables. Les épreuves externes auraient de forts impacts sur les contenus enseignés mais peu sur les pratiques des enseignants (Alderson & Wall, 1992; Cheng, 1999; Qi, 2005, 2007; Wall, 2000).

²⁸ Ce chercheur américain a réalisé une synthèse de 49 recherches qualitatives américaines qui traitent de l'effet des *high stakes tests* sur les pratiques des enseignants (ce qui représente 19 Etats, 96 écoles, 740 enseignants).

En se référant à ces diverses recherches, on peut conclure que l'effet de régulation des pratiques est très variable : une faible majorité d'enseignants serait touchée par cet effet. De plus, la variabilité de cet effet dans les classes est interrogée. Dumay et al. (2013) analysent ce phénomène dans leur étude de l'institutionnalisation d'instruments d'évaluation externe dans les établissements scolaires de CFB, en tant que « couplage/découplage entre le dispositif d'évaluation et les pratiques des enseignants ». Pour ce faire, référence est faite aux catégories proposées par Coburn (2004). Cette sociologue américaine a défini cinq catégories qui caractérisent l'intégration au non des nouvelles pratiques dans les classes en réponse aux pressions institutionnelles. Ces catégories viennent de son analyse des changements dans l'enseignement de la lecture en Californie.

- *Rejection* (rejet). L'enseignant ne reconnaissant pas ses croyances dans les nouvelles pratiques instituées, les refuse et maintient ses pratiques existantes.
- *Decoupling – Symbolic Response* (découplage - réponse symbolique). L'enseignant ne modifie pas ses pratiques, mais rend visibles certaines adaptations pour s'extraire des pressions institutionnelles.
- *Parallel structures* (structures parallèles). L'enseignant utilise deux types de pratiques : les nouvelles et les anciennes.
- *Assimilation* (assimilation). L'enseignant met en œuvre les nouvelles pratiques à sa manière, sans pour autant avoir compris leurs intentions pédagogiques et épistémologiques.
- *Accommodation* (accommodation). L'enseignant modifie entièrement ses pratiques en supprimant les anciennes.

Pour Coburn (2004), quatre facteurs peuvent influencer la réception des épreuves par les enseignants :

- *Degree of congruence* (congruence du message). La réponse des enseignants au message institutionnel dépend de son contenu, en particulier de sa congruence avec leurs pratiques et leurs croyances.
- *Degree of intensity* (intensité du message). La réponse des enseignants au message institutionnel dépend du degré d'opportunité que l'enseignant peut trouver dans celui-ci.
- *Degree of pervasiveness* (ampleur du message). La réponse des enseignants au message institutionnel dépend de l'importance de la réforme.
- *Degree of voluntariness* (imposition du message). La réponse des enseignants au message institutionnel dépend de sa nature : normative ou régulatrice. Dans le premier cas, on montre ce que l'enseignant « devrait » faire, ce qui encourage l'enseignant à faire un choix volontaire. Dans le deuxième cas, au contraire, on impose un changement par des règles et des sanctions.

Ce n'est donc pas simplement par des injonctions politiques que des changements sur le terrain sont réalisés. Ils dépendent de leur réception et de leur mise œuvre. Laveault (2012) propose trois éléments clés qui conditionnent l'implantation des politiques éducatives sur le terrain :

- *Alignement et cohésion interne.* Pour assurer une implantation des politiques, les pratiques doivent « s'aligner » sur des référentiels communs (tels que les plans d'études). Cet alignement garantit un enseignement uniformisé (les contenus et les pratiques). Pour assurer l'alignement de l'enseignement et de l'apprentissage, les systèmes éducatifs peuvent procéder de manière « *top-down* » (à l'aide d'épreuves externes) ou « *bottom-up* » (les enseignants au niveau local échangent sur leurs pratiques). Dans le premier cas, l'alignement dépend surtout de la qualité et de la validité des épreuves mais aussi du suivi des enseignants dans l'interprétation de leurs résultats, mais aussi des enjeux encourus. Dans le deuxième cas, la portée de l'alignement n'est que local.
- *Cohésion externe.* L'implantation des politiques éducatives doit aussi avoir une cohésion externe. Celle-ci dépend de plusieurs facteurs : les contenus doivent avoir une qualité et une validité scientifique (en lien avec les théories de de l'apprentissage et de l'enseignement). En effet, pour éviter toutes formes de résistances, les réformes éducatives doivent répondre à des critères de crédibilité de la part des enseignants, mais aussi de leur capacité à les mettre en œuvre (les changements doivent se situer dans leur zone de développement proximal). Enfin, les conséquences de cette implantation doivent permettre des impacts positifs et éviter des effets néfastes.
- *Intégration dans un cadre général.* Pour assurer une implantation des politiques, les nouvelles pratiques doivent s'intégrer dans un cadre général de pratiques préexistantes.

Laveault (2012b) cite notamment la réussite de l'intégration des politiques d'évaluation formative dans les pratiques des enseignants. Il représente cette implantation dans la figure suivante.

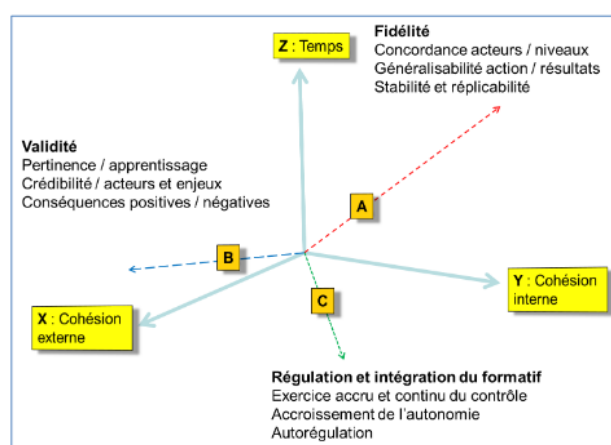


FIGURE 6 – Dimensions de mise en œuvre d'un cadre évaluatif (Laveault, 2009, p.5)

Finale­ment, la ques­tion de la ré­gulation des pra­tiques des ensei­gnants par l'é­valua­tion ex­ter­ne reste ou­verte. Le pro­ces­sus par le­quel il se pro­duit reste in­expli­qué. Il est dif­fi­cile de plani­fier un tel chan­ge­ment. Le chan­ge­ment est pos­sible s'il est pro­gres­sif et ac­ces­sible aux ac­teurs de ter­rain et qu'il n'en­traîne de fait pas (ou peu) de ré­sis­tances (Carette & Dupriez, 2009). Tou­te­fois, la ré­gulation des pra­tiques com­porte des ef­fets in­at­ten­dus et né­fas­tes à l'appren­ti­ssage des élève­s. Il s'agit des ef­fets de *teaching to the test* ou de ré­trécis­se­ment du cur­ricu­lum²⁹. Ces ré­sul­tats met­tent alors en doute le fait que l'é­valua­tion ex­ter­ne puisse être un fac­teur de ré­gulation ef­fi­cace, sur­tout sans au­tre me­sure d'ac­com­pag­ne­ment. Cet échec tient à des fac­teurs qui sont pro­pres au con­texte, au dis­po­si­tif d'é­valua­tion et aux ensei­gnants.

Aussi, les ef­fets de ré­gulation de l'é­valua­tion ex­ter­ne varient beau­coup selon les di­men­sions des pra­tiques ensei­gnantes qu'ils tou­chent. Spratt (2005)³⁰ dé­montre que cet ef­fet peut tou­cher de nom­breux as­pects dans la classe : le cur­ricu­lum, le ma­té­riel pé­da­go­gique, les at­ti­tudes et les sen­ti­ments des élève­s et des ensei­gnants et l'appren­ti­ssage des élève­s. Cattonar, Dumay, Mangez, & Maroy (2010)³¹ iden­ti­fient 6 di­men­sions de pra­tiques tou­chées (ils in­di­quent le nom­bre d'ensei­gnants con­cernés) : é­valua­tion in­ter­ne (23/28); pra­tiques (17/28); pré­pa­ra­tion (12/28) ; plani­fi­ca­tion des ob­jectifs (9/28) ; ry­thme (5/28) ; ré­uti­li­sa­tion sup­port (5/28).

Nous dé­tail­lons les ef­fets de ré­gulation de l'é­valua­tion ex­ter­ne sur les pra­tiques des ensei­gnants sur deux axes : leurs pra­tiques d'é­valua­tion in­ter­ne et leurs pra­tiques d'ensei­gne­ment.

a) Ré­gulation des pra­tiques d'é­valua­tion in­ter­ne

Parmi les pra­tiques des ensei­gnants, c'est sur­tout sur les pra­tiques d'é­valua­tion de l'ensei­gnant dans sa classe (in­ter­ne) que les cher­cheurs iden­ti­fient les ef­fets de l'é­valua­tion ex­ter­ne. Voici des ré­sul­tats d'en­quête­s em­pi­ri­ques :

- Pour Cattonar et al. (2010), c'est sur l'é­valua­tion in­ter­ne que les ef­fets des é­preu­ves ex­ter­nes sont les plus mar­qués (23 ensei­gnants sur 28). L'é­valua­tion ex­ter­ne est pour eux un ou­til com­plé­men­taire qui peut les aider dans l'é­valua­tion des appren­ti­ssages de leurs élève­s. La ma­jeure partie des ensei­gnants in­ter­ro­gés di­sent mo­di­fier leurs pra­tiques d'é­valua­tion par rap­port à l'ex­em­ple de l'é­valua­tion ex­ter­ne en dif­fé­rents points : imi­ta­tions des con­di­tions de pas­sa­tion (voire un coaching) ; mo­di­fi­ca­tion de la fa­çon de po­ser les ques­tions (sans tou­te­fois

²⁹ Ces ef­fets sont l'ob­jet des points 6.2.4 (p.66-67) et 6.2.5 (p.68).

³⁰ L'auteure pro­pose une re­vue des ar­ti­cles traitant de l'ef­fet de *washback*.

³¹ Les auteurs ont in­ter­ro­gé 28 ensei­gnants bel­ges (CFB) du degré pri­maire au su­jet de leur ré­cep­tion et leur usage des é­preu­ves ex­ter­nes.

remettre en cause ce qui est évalué) ; évaluation d'autres processus cognitifs (notamment les compétences). Des variations dans ces effets sont observées selon la fonction des épreuves externes (les épreuves non-certificatives influencent davantage les pratiques évaluatives des enseignants) mais aussi selon le degré d'enseignement (les enseignants des grandes classes sont davantage influencés). Ces variations ne peuvent pourtant pas être généralisées, elles ne font pas l'objet d'analyse statistiques.

- En France, selon une étude de la DEP (2005) au niveau du collège, la moitié des enseignants se disent influencés par les épreuves diagnostiques nationales dans leurs pratiques d'évaluation interne (Pons, 2011).
- L'étude de Longchamp (2010) sur l'utilisation des épreuves cantonales (EC) de mathématiques de 6^{ème} année par les enseignants vaudois nous renseigne également sur l'impact de celles-ci sur leurs pratiques d'évaluation : dans l'ensemble, ces épreuves n'influencent que peu les pratiques évaluatives des enseignants. Sur les 20 enseignants de mathématiques interrogés, 17 affirment que ni le contenu ni la forme des EC (plutôt axées sur les compétences et la résolution de problèmes) n'influencent leurs pratiques d'évaluation. Plusieurs raisons sont avancées : leurs pratiques se rapprochent déjà de celles de l'EC (7 enseignants sur 20) ; ils modifient en partie leurs pratiques tout en conservant des éléments de leurs pratiques pré-existantes (4/20) ; ils conservent leurs propres pratiques (6/20). Pour les enseignants qui affirment modifier leurs pratiques selon l'EC (3/20), ces changements touchent surtout les consignes (consignes davantage claires) et les contenus (davantage d'objectifs touchés et inclusion d'objectifs à long terme). Les résultats des élèves aux EC n'influencent qu'une minorité d'entre eux (3/20) qui disent modifier essentiellement leurs barèmes.
- L'étude de Ntamakiliro et Tessaro (2002) démontre que les enseignants genevois du degré primaire prennent l'évaluation externe comme un moyen utile pour « renouveler ou développer leurs outils d'évaluation interne ».
- L'étude de Dierendonck (2008)³² montre que si les enseignants suisses romands du degré primaire perçoivent les épreuves cantonales comme utiles, c'est surtout car elles leur fournissent de nouvelles idées sur la manière d'évaluer les apprentissages de leurs élèves.

³² L'auteur a interrogé des enseignants du degré primaire de différents cantons suisses romands (Neuchâtel, Vaud et Fribourg) par *focus group* pour investiguer leurs perceptions de 6 épreuves externes (chaque groupe étant questionné sur une épreuve).

L'évaluation externe peut ainsi fournir des pistes aux enseignants pour développer leurs pratiques d'évaluation. C'est notamment le cas pour les pratiques d'évaluation des compétences. Dans la plupart des pays européens et en Suisse, les plans d'études sont aujourd'hui formulés sous forme de compétences à atteindre par les élèves (Eurydice, 2009; Mons & Pons, 2006). De ce fait, les pratiques évaluatives doivent évoluer. Les enseignants dans leur classe (évaluation interne) et les administrations (évaluation externe) doivent évaluer l'atteinte de ces compétences inscrites dans les référentiels (Carette, 2009; Mottier Lopez & Laveault, 2008). Pourtant, certains chercheurs mettent en doute cet effet de régulation. L'opérationnalisation de telles épreuves externes serait difficile et elles seraient trop distantes des pratiques des enseignants en classe. De plus, la plupart des épreuves externes ne testeraient pas réellement des compétences. Enfin, certains s'interrogent sur le fait que les compétences soient un indicateur pertinent pour évaluer la performance d'un système éducatif (Behrens, 2005; Carette, 2007; Crahay, 2009; Rey, 2005). La logique des compétences rend encore plus complexe l'activité évaluative (Crahay, 2009). Le message (l'attendu) doit être clair. Carette (2009) montre que certaines évaluations voulant évaluer des compétences ne le font pas vraiment. Les évaluations externes sensées évaluer des compétences sont en fait plutôt similaire à des évaluations critériées et proches de la pédagogie par objectifs (Carette & Dupriez, 2009). Les enseignants risquent alors de mal interpréter ce qui est attendu d'eux.

Les enseignants face aux messages et aux injonctions contradictoires risquent d'interpréter les objectifs de la réforme à leur manière au regard de leurs croyances (Carette, 2009, p.156).

Carette et Dupriez (2009) relèvent que la double fonction (certification et régulation des pratiques) qu'il est parfois donné aux évaluations externes « brouille le message ». Les évaluations certificatives ne peuvent pas être empruntées de pédagogies innovantes (notamment pour l'approche par compétences), devant faire surtout office de moyen de comparaison des élèves.

Pour Carette (2009), outre l'imposition d'une norme par les évaluations externes, une autre possibilité serait d'éviter les « discours pédagogiques prescrits » (*top-down*) et d'encourager les enseignants à réfléchir sur les nouvelles notions, de les rendre acteurs de ce changement (*bottom-up*). La proposition de Lafontaine (2012) va dans l'autre sens : les contenus des épreuves externes devraient se rapprocher des pratiques effectives des enseignants. Elle considère « inadmissible » pour les épreuves certificatives et « inadéquat » pour celles non-certificatives. Les épreuves externes non-certificatives pourraient servir à développer les compétences des élèves, mais ne devraient alors pas être trop innovantes, elles devraient se situer dans « la zone de développement proche des enseignants ».

Toutefois, les épreuves externes peuvent également être le vecteur de « mauvaises » pratiques évaluatives. En effet, les concepteurs d'épreuves sont souvent contraints de limiter leur questionnement à des niveaux taxonomiques bas, pour des raisons pratiques et psychométriques (correction rapide et objective, passation facilitée pour les enquêtes à large échelle). Souvent, surtout

dans les tests commercialisés et/ou internationaux, les questions à choix multiples sont privilégiées. Ces exemples peuvent être reproduits par les enseignants, qui souvent limitent leur questionnement à de simples restitutions de connaissances. De plus, ces tests sont pour la plupart construits dans un objectif de sélection normative des élèves, cherchant à obtenir une distribution gaussienne des résultats. L'évaluation externe (qui devrait pourtant être un modèle pour les enseignants) se retrouve alors en porte-à-faux avec les théories et les didactiques actuellement préconisées (Davaud & Cardinet, 1992; Madaus & Kellaghan, 1992; Monseur & Demeuse, 2005).

b) Régulation des pratiques d'enseignement

Les épreuves externes peuvent également avoir des effets sur les pratiques d'enseignement. Ces effets ont été identifiés dans plusieurs contextes grâce à des méthodologies diverses. Plusieurs dimensions de l'acte d'enseigner sont touchées, nous les détaillons ici.

- *Matériel d'enseignement.* Par le biais d'épreuves externes, les enseignants peuvent être amenés à modifier leurs supports didactiques. C'est notamment l'un des effets observés en CFB, qui reste toutefois marginal (5 enseignants interrogés sur 28) (Cattonar et al., 2010). Les enseignants sont par exemple amenés à utiliser d'anciennes épreuves externes (tout ou en partie) dans leur enseignement ou à s'en inspirer pour créer du matériel de cours (Longchamp, 2010). Cet effet est également observé dans la littérature sur le *washback* (Smith, 1991; Spratt, 2005).
- *Méthodes d'enseignement.* Il se peut que les évaluations externes conduisent à des changements dans les méthodes d'enseignement (compris dans le sens de la façon dont l'enseignant amène ses élèves à apprendre). D'un côté, il se peut que les épreuves externes conduisent les enseignants vers des pratiques nouvelles en adéquation avec les théories sur l'apprentissage et les plans d'études. En effet, l'étude de Cattonar et al. (2010) montre qu'en CFB, les épreuves externes ont mené les plupart des enseignants interrogés vers « une plus grande application de la pédagogie des compétences » (23/28). D'un autre côté, si les enseignants sont contraints à suivre un curriculum serré et strict, avec peu de temps pour le faire, il y a un risque que leur créativité et leur initiatives soient inhibées. Ainsi, les enseignants sont contraints dans leur choix et dans la variété des contenus et des pratiques (Monseur & Demeuse, 2005). L'esprit d'innovation est donc bridé, surtout si la qualité des tests est mauvaise et si les curricula sur lesquels ils s'alignent ne promeuvent pas des contenus ou des méthodes innovants (Looney, 2009). Les pratiques des enseignants peuvent alors aller à l'encontre des théories pédagogiques constructivistes modernes (qui sont pourtant valorisées dans les formations, les programmes et matériaux didactiques) se concentrant sur des compétences basiques (*basic skills*). (Darling-Hammond, 1994).

Si un curriculum est fixé et qu'un système d'évaluation externe limite les possibilités de contextualisation, la connaissance risque de ne plus être construite et les approches interdisciplinaires ne seront plus prisées (Monseur et Demeuse, 2005, p.505).

Les effets de l'évaluation externe peuvent également être néfastes lorsque les enseignants veulent faire réussir à tout prix leurs élèves, il s'agit alors de pratiques de bachotage (*teaching to the test*)³³. Les enquêtes issues de la littérature sur le *washback* sont à nouveau partagées sur ce sujet. Si les résultats de certaines études montrent que les tests standardisés de langue n'ont pas d'influence sur la façon dont les enseignants mènent leurs cours, d'autres montrent que les enseignants changent leurs méthodes de manière positive ou non (Smith, 1991; Spratt, 2005).

- *Planification de l'apprentissage.* Une partie des enseignants interrogés par Cattonar et al. (2010) identifient également des effets dus aux épreuves externes sur leurs pratiques de planification des apprentissages (9/28). Ils peuvent être amenés à programmer les objectifs en fonction de l'évaluation externe. La plupart des enseignants vaudois interrogés dans l'étude de Longchamp (2010) affirment planifier les apprentissages de leurs élèves en fonction de l'épreuve cantonale (2/28).
- *Accompagnement des élèves dans leurs apprentissages.* L'évaluation externe peut également amener les enseignants à modifier la manière dont les enseignants accompagnent leurs élèves dans leurs apprentissages. Les épreuves externes peuvent en effet les amener à ne pas pouvoir respecter le rythme des élèves. Par exemple, la plupart des enseignants vaudois interrogés dans l'étude de Longchamp (2010) disent devoir « veiller à conserver un certain rythme, voire l'accélérer » pour réussir à traiter tous les contenus annoncés dans l'épreuve cantonale. De plus, les épreuves externes dont la finalité est normative (par exemple épreuve d'orientation) et/ou à fort enjeux peuvent amener les enseignants à s'occuper en priorité de certains profils d'élèves (Mons, 2009). En analysant le contexte scolaire texan, Booher-Jennings (2005) démontre que les enseignants, du fait des tests à hauts enjeux, tendent à se concentrer principalement sur les élèves « moyens » qui ont le plus de chance d'influencer les résultats du test d'une classe ou d'un établissement.

6.2.4. Teaching to the test – bachotage

Comme nous l'avons exposé précédemment, l'évaluation externe peut être vectrice de formation pour les enseignants, en faisant évoluer leurs pratiques d'enseignement et d'évaluation. Cependant, de nombreuses études montrent que ces dispositifs d'évaluation, surtout quand ils sont emprunts d'enjeux forts, peuvent également être un vecteur de pratiques inattendues et néfastes. Il peut s'agir du

³³ Cet effet est l'objet du point 6.2.4 (p.66-68).

bachotage (en référence au baccalauréat français) ou *teaching to the test* : les enseignants adoptent des pratiques d'entraînement intensif (*coaching*) pour préparer leurs élèves à réussir les épreuves externes plutôt qu'à développer les apprentissages inscrits dans les plans d'études (Looney, 2009; Mons, 2009). Un moyen notamment souvent utilisé est de faire passer des examens des années antérieures ou des exercices qui y ressemblent fortement (Laveault, 2009b; Monseur & Demeuse, 2005).

Dans les pires scénarios cependant, ces épreuves sont considérées au premier niveau et l'enseignement se concentre sur la réussite du test et non sur celle des attentes du programme (« teach to test ») (Laveault, 2009b, p. 11).

La plupart des recherches qualitatives américaines sur les effets des *high stakes tests* sur les pratiques des enseignants montrent que les enseignants sont amenés au *teaching to the test* (24 études sur 41) et à des méthodes d'enseignement davantage transmissives (*teacher-centered*) (32 études sur 41) (Au, 2007). Un rapport sur les épreuves cantonales du canton de Vaud relève également ces pratiques (Ntamakiliro & Moreau, 2011). Ces pratiques s'accroissent plus le test est proche (Spratt, 2005) et plus ses enjeux sont importants (Mons, 2009). Dans ce cas, il a été observé que les enseignants passent beaucoup de temps à « entraîner » leurs élèves à réussir les tests (B. Jones, 2007).

Cet effet ne crée pas de véritable débat dans la communauté scientifique. La (très) grande majorité des chercheurs dénoncent ce phénomène (Bélaïr, 2005; Darling-Hammond, 1994; Haney, 2000; Jones et al., 1999). Ces exercices d'entraînement ne sont ni pertinents ni efficaces pour plusieurs raisons. Premièrement, les enseignants qui font du *teaching to the test* préparent leurs élèves à réussir un examen mais pas à maîtriser les objectifs des programmes. Ceci crée donc un rétrécissement du curriculum enseigné et l'alignement des contenus n'est pas respecté. Deuxièmement, cet entraînement intensif ne permet pas un réel apprentissage de la part des élèves. On prépare les élèves à répondre à des questions types, sans cultiver leurs compétences, pourtant visées par les examens. Les objectifs poursuivis sont donc d'un niveau taxonomique très bas. Ces pratiques pédagogiques plutôt transmissives mettent les élèves dans un rôle trop passif.

Elles (les pratiques de *teaching to the test*) se basent sur une conception désuète de l'intelligence et ne les (les élèves) incitent pas à s'engager dans la structuration de tâches, la production d'idées nouvelles, la synthèse et la résolution de problèmes » (Monseur & Demeuse, 2005, p.508).

Troisièmement, à cause de ces pratiques, les résultats du test ne reflètent pas les performances réelles de l'élève, ils sont trop élevés. On appelle ce phénomène *score inflation* (Koretz, 2005). Ceci a comme conséquence de ne pas pouvoir interpréter correctement les résultats des épreuves. Enfin, ces pratiques renforcent « la ségrégation scolaire », desservant les élèves les plus faibles (Normand, 2005).

Néanmoins, Popham (2001) trouve toutefois certaines vertus à cet effet : il pousse les enseignants à aligner les contenus enseignés sur le programme et, sous sa forme de *coaching*, à améliorer les performances des élèves. Cependant, pour l'auteur, des conditions doivent être respectées :

l'enseignant doit se servir des items des épreuves comme moyen de développer des apprentissages (*curriculum-teaching*) et non « cloner » ces items pour préparer les élèves à y répondre (*item-teaching*).

6.2.5. Rétrécissement du curriculum enseigné

L'effet de rétrécissement du curriculum (*narrowing of the curriculum*) est identifié par la littérature, comme un effet inattendu et indésirable. D'une part, il peut s'agir, dans les cas où l'enseignant est généraliste (il enseigne toutes les branches dans une seule classe), d'une plus grande attention (en termes d'heures et d'importance) données aux disciplines évaluées dans l'épreuve externe (généralement mathématiques et langues) au détriment des branches "secondaires" (arts, sport...). D'autre part, l'enseignant est amené à enseigner des compétences plus basiques (objectifs plutôt cognitifs et mesurables facilement) qui sont le plus souvent évalués dans les tests à grande échelle au détriment de compétences plus complexes qui font pourtant partie des exigences des référentiels (créativité, autonomie...). Au final, c'est l'épreuve externe qui détermine le curriculum et non l'inverse (B. Jones, 2007; Mons, 2009; Monseur & Demeuse, 2005).

Cet effet a été identifié par de nombreuses recherches dans différents contextes, comme le montre la métasynthèse de Au (2007). La plupart des recherches analysées identifient cet effet (34 recherches sur 41), surtout lorsque les épreuves externes comportent des enjeux importants (M. Gail Jones, Jones, & Hargrove, 2003; Madaus, 1991). Cet effet est également démontré par les études empiriques sur le *washback effect* (Spratt, 2005).

L'effet de rétrécissement du curriculum peut s'apparenter à celui de *teaching to the test* (ou en faire partie). Généralement, dans la littérature, ces deux effets sont traités séparément. Cependant, la métasynthèse de Au (2007) montre que les effets de *teaching to the test* et de rétrécissement du curriculum vont souvent de pair dans le contexte américain (26 études sur 41).

6.2.6. Développement du professionnalisme des enseignants

Comme exposé dans le point 5.2.3 (p.37), dans un système scolaire décentralisé où l'enseignant et l'établissement ont gagné en autonomie, l'évaluation standardisée des acquis des élèves est un outil utile aux décideurs pour « responsabiliser » les acteurs de terrain (Mons, 2009). Alors qu'auparavant l'institution était garante de la qualité du système scolaire, ce sont les enseignants qui sont aujourd'hui tenus responsables des résultats de leurs élèves et, de fait, de la qualité du système tout entier (Mons & Dupriez, 2010). Néanmoins, aujourd'hui, les compétences et l'autonomie des enseignants sont remises en cause par la société, la confiance se « lézarde » (Maroy, 2014). Dans ce sens, l'évaluation externe

est devenue un outil de contrôle « à distance » du travail des enseignants (Mons, 2009) et de leur « professionnalisme » (Klieme et al., 2004; Mons & Dupriez, 2010).

Les enseignants, rendus responsables des résultats des élèves devant leur supérieur hiérarchique et la société civile par la publication des performances de leur école, travailleraient davantage à la réussite des élèves et amélioreraient leur professionnalisme aussi bien par des formations que des échanges entre collègues au sein des écoles, les résultats aux tests donnant lieu à des analyses réflexives dans les établissements (Mons & Dupriez, 2010, p. 48).

Les enseignants ont aujourd'hui une « obligation morale » de se préoccuper des résultats de leurs élèves « doublée d'une incitation cognitive et morale à la réflexivité » (Maroy, 2013, p. 21). Sous cette forme de « nouvelle responsabilisation » des acteurs de terrain, les résultats des épreuves externes permettent aux décideurs de contrôler et de développer le « professionnalisme » des enseignants. Ceux-ci, si l'on s'appuie sur une définition du professionnalisme « managérial ou instrumental »³⁴, sont considérés comme des « exécutants qualifiés et autonomes » (Maroy, 2014):

- *Qualifiés*. On reconnaît à l'enseignant une expertise dans son travail (scientifique et/ou expérience). Il maîtrise les moyens qui lui permettent de résoudre les problèmes ;
- *Autonomes*. L'enseignant ne décide pas des objectifs et des orientations du travail. Ils sont normatifs, venant des décisions des autorités. L'enseignant est toutefois autonome dans le choix des moyens à utiliser.

L'enseignant doit donc prouver son « professionnalisme » en atteignant les objectifs qui lui sont assignés, les résultats des épreuves externes en étant la mesure, une évaluation de son « professionnalisme ». Aujourd'hui, le « professionnalisme » attendu des enseignants comprend plusieurs dimensions (Maroy & Cattonar, 2002) :

- *Spécialiste de l'apprentissage*. L'enseignant ne peut aujourd'hui plus se contenter de maîtriser et de transmettre des savoirs. Il doit surtout « se centrer sur le processus d'apprentissage des élèves » dans un modèle d'enseignement constructiviste et différencié ;
- *Praticien réflexif*. L'enseignant ne peut aujourd'hui plus se contenter d'appliquer des « recettes préétablies ». Il doit être un « praticien réflexif », capable de s'adapter à toutes les situations d'enseignement, de manière autonome, en analysant ses pratiques et les résultats de ses actions ;

³⁴ Au contraire, le professionnalisme « vocationnel ou démocratique » comprend une certaine liberté des acteurs quant à l'acceptation et/ou à la négociation des objectifs et des orientations. Vu sous cet angle, la régulation par les résultats ne permet pas le développement du professionnalisme des enseignants.

- *Travail en équipe*. L'enseignant doit travailler dans une équipe ancrée dans un établissement (voir point précédent).

Néanmoins, le fait que les enseignants soient capables de « lire » les résultats d'évaluations externes puis de réguler leurs pratiques en fonction de ceux-ci est mis en doute. Il est constaté que cette intention reste de « nature instrumentale et devient une voie de diffusion des pratiques éducatives souhaitées par les autorités plutôt qu'une voie de *d'empowerment* des professionnels » (Mons & Dupriez, 2010, p. 57). Aussi, ces auteurs constatent que lorsque les dispositifs de *testing* sont de nature réflexive et dont les enjeux sont faibles, les enseignants ne s'engagent que peu dans le développement de leur professionnalisme.

Cet effet de «développement du professionnalisme des enseignants» couplé à une responsabilisation des enseignants est identifié par quelques recherches. Cattonar et al. (2010) démontrent que les enseignants, par le biais des épreuves externes, ont l'impression d'être jugés responsables des résultats des élèves, surtout lorsqu'ils sont confrontés à des épreuves de fin de cycle. Cette responsabilisation peut être accentuée par l'attitude des inspecteurs ou par la comparaison entre collègues et est vécue de manière négative et injuste ; on oublie en effet qu'au moment de l'évaluation, l'élève a toute une carrière scolaire derrière lui et qu'il a eu avant cela plusieurs enseignants. La recherche montre que les enseignants attribuent généralement la responsabilité des résultats de leurs élèves à leurs propres pratiques, plus rarement aux caractéristiques des épreuves ou aux caractéristiques de leur classe. Les résultats des épreuves les amènent alors à remettre en cause leurs pratiques et à chercher à les améliorer. Dans le champ de la linguistique également, certaines recherches démontrent que les enseignants, se sentant responsables de la réussite de leurs élèves aux tests standardisés, prêtent davantage d'attention à leurs pratiques (Spratt, 2005).

6.2.7. Collaboration – travail en équipe

Dans un système décentralisé où les établissements jouissent d'une grande autonomie, un effet qui peut être attendu des évaluations externes par les autorités est celui de pousser les enseignants à collaborer entre eux. Aujourd'hui dans les systèmes scolaires, pour améliorer la qualité de l'enseignement, il est demandé aux enseignants de travailler en équipe afin de rompre avec « l'isolement actuel dans lequel travaillent (certains) enseignants » (Maroy & Cattonar, 2002, p. 6). Les enseignants, dont les rôles se sont complexifiés, sont encouragés à s'investir dans la gestion de l'établissement et à développer des pratiques communes (notamment dans les projets ou les conseils d'établissement). Les épreuves externes sont donc un moyen pour les décideurs d'amener les enseignants à travailler en équipe de manières diverses (Demailly, 2001; Maroy & Cattonar, 2002; Wirthner & Ntamakiliro, 2008) :

- Echanger entre eux sur les résultats de leurs élèves aux épreuves externes ;
- Echanger entre eux sur les critères d'évaluation des épreuves externes ;
- Echanger entre eux et travailler ensemble sur la régulation de leurs pratiques ;
- Planifier ensemble les apprentissages des élèves...

Cet effet est notamment observé par des recherches en Belgique (Cattonar et al., 2010), en Angleterre (Maroy, 2009) ou en Suisse (Longchamp, 2010). Toutefois, dans tous ces contextes, cette collaboration peut être parfois vécue par les enseignants comme une obligation institutionnelle.

6.2.8. *Déprofessionnalisation – démotivation des enseignants*

Au contraire des éléments présentés au point précédent, les épreuves externes, inscrites dans des politiques de responsabilisation des enseignants, peuvent avoir des effets néfastes sur « l'engagement et le professionnalisme » des enseignants (Maroy, 2012a). Le stress engendré par les résultats, mais aussi la multiplication des mesures, peuvent entraîner une certaine démotivation.

En particulier, dans certains cas, la perception du métier devient négative, la satisfaction dans l'activité professionnelle diminue tandis que la perception d'un stress nouveau apparaît (Mons, 2009, p. 124).

Cette démotivation a notamment été démontrée par des études américaines et tout particulièrement dans les établissements défavorisés (B. Jones, 2007). Les nouvelles politiques de *testing* peuvent amener les enseignants vers un sentiment de « déprofessionnalisation » de leur métier (Maroy & Cattonar, 2002). Ce phénomène est observé par des recherches dans de nombreux pays européens notamment en France, en Angleterre, au Danemark (Osborn, 2006), ou en CFB (Dupriez, 2004; Maroy & Cattonar, 2002). Plusieurs raisons sont avancées :

- a) *Pressions accrues*. La responsabilisation des enseignants par le biais du *testing* (le fait d'être surveillé et la nécessité de devoir prouver ses compétences) accroît les pressions qui reposent sur les enseignants (Maroy, 2009).
- b) *Perte de valeurs*. Dans un cadre politique basé sur la performance et les résultats, certains enseignants ne reconnaissent plus les valeurs de leur métier. Des éléments importants tels que l'investissement personnel ou la créativité sont rendus difficiles, voire impossibles, en raison de la pression liée à la performance. Les enseignants peuvent avoir l'impression d'être devenus des « techniciens » qui ne font que transmettre des connaissances prédéfinies par des autorités externes (Osborn, 2006).
- c) *Perte de confiance*. Ces politiques éducatives basées sur les résultats ont notamment été instaurées dans une période de méfiance quant aux qualités professionnelles des

enseignants (engagement, compétences professionnelles). De ce fait, l'institution scolaire et les enseignants ne sont plus de facto « dignes de confiance » aux yeux de la société et des autorités, leur professionnalisme devant être surveillé par des moyens externes (Maroy, 2012a).

- d) *Perte d'autonomie*. Les nouvelles politiques éducatives sont généralement basées sur une grande autonomie des établissements et des enseignants, mais paradoxalement, elles s'appuient également sur un important « contrôle à distance » par le biais du *testing*. Dans ces nouveaux modes de régulation, les enseignants ont donc gagné une certaine liberté de moyens, mais on leur impose une responsabilité de résultats. Leur autonomie est alors toute relative et les enseignants « ne sont plus totalement seuls pilotes dans leurs classes » (Mons, 2008a). Dans ce sens, le professionnalisme des enseignants est remis en cause (considéré comme vocationnel ou démocratique) ou limité (considéré comme instrumental):

Il n'y a menace sur le professionnalisme enseignant que si, dans le même temps, on réduit trop son autonomie sur les moyens ou son autonomie de réflexion sur sa pratique. Dans ce cas de figure, l'enseignant deviendrait en effet un simple « exécutant qualifié », si par exemple on combinait un fort développement de l'évaluation des acquis des élèves et une prescription des pratiques enseignantes (Maroy, 2014, p.211).

6.2.9. *Tricheries*

Un effet peu louable des épreuves externes sur les pratiques des enseignants peut être de les amener à tricher. Comme déjà exposé, les résultats de certains tests standardisés peuvent être entachés d'irrégularité dans les échantillons, les élèves en difficultés en étant exclus. Mais les épreuves, en cas de forts enjeux, peuvent aussi entraîner des tricheries dans les classes (Bélair, 2005).

6.3. Les effets de l'évaluation externe sur les cadres intermédiaires

L'évaluation externe a des effets sur les acteurs du niveau intermédiaire (mésos) de la scolarité: directeurs d'établissement, inspecteurs, administrations locales... Les recherches qui en traitent sont peu développées. Toutefois, elles montrent que ces acteurs sont friands de telles mesures, davantage encore que les enseignants. L'évaluation externe leur donne des outils utiles à leurs prises de décisions quant au suivi des enseignants et aux performances des élèves (Mons, 2009). Le *testing* a créé de nouvelles opportunités de communication entre enseignants et cadres intermédiaires. Cet « espace de

dialogue » permettrait d'améliorer le système tout entier en évitant les tensions entre l'intérêt individuel et général (Crahay, 2009).

Cependant, outre leurs raisons pédagogiques, les évaluations externes seraient davantage utilisées par les cadres intermédiaires afin d'asseoir leur autorité sur les enseignants et pour se créer une bonne image, notamment par la compétition entre établissements (Pons, 2011). Finalement, paradoxalement aux visées d'autonomie prônées par les nouveaux types de régulation post-bureaucratiques, l'évaluation externe a donc renouvelé et renforcé le rôle des cadres intermédiaires. Dans certains contextes, notamment en France, les cadres intermédiaires vont même jusqu'à créer leurs propres outils d'évaluation externe locaux. Aussi, l'évaluation standardisée a fait naître de nouveaux acteurs intermédiaires et de nouvelles structures tels que les agences de *testing*, les responsables d'épreuves, les statisticiens, etc. (Mons, 2009).

La recherche de Cattonar et al. (2010) illustrent bien l'effet des épreuves externes sur la relation entre enseignants et les cadres intermédiaires. Les enseignants interrogés témoignent du fait que les directions d'établissement et les inspecteurs, à des degrés variables, utilisent les résultats des épreuves pour légitimer leurs interventions et pour mettre une certaine pression sur les enseignants.

6.4. Les effets de l'évaluation externe sur les élèves

L'évaluation externe des acquis des élèves entraîne des effets sur les élèves. Ils touchent surtout leur motivation. Certes, les résultats de tels dispositifs peuvent permettre à l'élève de savoir où il se situe par rapport à une norme. Toutefois, cette information serait bénéfique pour les « bons » élèves, mais néfaste pour les élèves en difficultés. De plus, cette valeur informative n'aurait que peu d'effet, l'élève n'étant pas en mesure d'analyser ses résultats et de réguler ses apprentissages en fonction de ceux-ci. Bien souvent, dans le cas d'évaluation externe, l'enseignant ne remplit pas non plus ce rôle de médiateur (Crahay, 2009).

Pourtant, dans certain contextes, les dispositifs d'évaluation externe sont mis en place avec l'intention de motiver les élèves (du moins extrinsèque) et donc d'augmenter leurs performances. Les études qui se sont penchées sur le sujet présentent des résultats contradictoires (Looney, 2009; Mons, 2009). Certaines études américaines, qualitatives ou quantitatives, montrent que les dispositifs de *testing* notamment à forts enjeux réussissent à motiver les élèves et à améliorer leurs performances (Bransford, Brown, & Cocking, 2000; Cizek, 2001; Debard & Kubow, 2002). Toutefois, d'autres études démontrent le contraire. Les tests externes feraient baisser la motivation intrinsèque des élèves pour l'apprentissage et l'école, les pratiques pédagogiques en lien étant moins stimulantes (M. Jones et al., 1999; M. Jones et al., 2003). Surtout, ils seraient la source d'un nouveau stress. Cette forme d'anxiété vis-à-vis de l'école est néfaste à l'apprentissage, allant même jusqu'à la dépression de

certaines élèves en cas de grande pression (notamment au Japon ou en Corée). Justement, certaines études montrent que les élèves les moins anxieux face aux tests sont ceux qui les réussissent le mieux (Looney, 2009). Enfin, le stress face aux tests standardisés serait la cause de nombreux abandons, notamment dus au manque de soutien des enseignants pour les élèves en difficulté (Amrein-Beardsley & Berliner, 2003; Haney, 2000). L'étude de Cattonar et al. (2010) illustre ces propos. Pour les enseignants interrogés, les épreuves externes sont source de stress pour les élèves. Alors que les enseignants essaient plutôt de dédramatiser ces situations, les parents amplifieraient souvent cette anxiété.

6.5. Les effets de l'évaluation externe sur les parents

Les évaluations externes touchent également les parents des élèves. Elles leur permettent tout d'abord de s'informer du niveau de performance de leur enfant, bien souvent prises comme une mesure davantage objective que l'évaluation de l'enseignant (Crahay, 2009). En effet, les parents sont généralement très majoritairement demandeurs de telles épreuves, reconnaissant toutefois leurs effets négatifs (Mons, 2009). Les études sur le sujet sont très rares.

Les épreuves externes poussent bien souvent les parents à des stratégies tout à fait louables et compréhensibles mais qui sont au final inévitables. Les parents qui ont le plus de moyens financiers, à l'approche de ces examens, des cours d'appui, des cours particuliers. Ces moyens ont certainement des effets positifs pour la réussite des évaluations externes, mais peuvent également avoir des effets négatifs sur l'apprenant, surchargé de travail. Cet effet est notamment observable au Japon, où il contribuerait au taux de suicide élevé chez les jeunes. Aussi, ce type d'aide ne peut qu'être proposé par des classes socioéconomiques élevées, ce qui crée une disparité sociale entre les élèves, alors que l'école devrait être capable d'offrir les mêmes chances à tous (Monseur & Demeuse, 2005).

Cattonar et al. (2010) montrent que l'attitude des parents vis-à-vis des épreuves externes peut varier. Elle peut aller de l'indifférence à une grande anxiété. Les enseignants interrogés soulignent que les parents anxieux transmettent souvent ce stress à leurs enfants et mettent une grosse pression sur les enseignants. Afin d'éviter ce dernier cas, les enseignants disent éviter de trop communiquer au sujet des épreuves.

7. Facteurs qui influencent les effets de l'évaluation externe sur les pratiques des enseignants

Comme nous l'avons démontré, la littérature identifie de nombreux effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants, notamment par des recherches empiriques. Toutefois, rares sont celles qui explorent les processus par lesquels se produisent ces effets, leur variabilité et les facteurs qui l'influencent. Les enjeux que comportent les différents dispositifs sont souvent relevés comme l'unique facteur marquant une différence entre les effets de l'évaluation externe. L'objectif de cette partie est justement d'exposer tous les facteurs, présents dans la littérature, qui peuvent expliquer la variabilité de ces effets sur les pratiques des enseignants.

Toutefois, si les évaluations semblent influencer les pratiques de classe, il est difficile de déterminer la nature de cette influence (Carette & Dupriez, 2009, p. 41).

C'est surtout dans la littérature du *washback effect* que les facteurs qui ont un impact sur les effets des épreuves externes sont investigués. Jusqu'aux années 2000, ce champ de recherche questionne surtout la présence ou non d'effets liés aux épreuves standardisées. Il s'attache également à positionner ceux-ci sur une dualité : effets positifs vs. effets négatifs. Depuis, les intérêts de recherche ont évolué. S'il est indéniable que ces effets (positifs ou négatifs) existent, ils sont de natures très complexes et leur variabilité va au-delà de cette dualité. La question essentielle est aujourd'hui de savoir pourquoi telle épreuve externe entraîne tel(s) effet(s) (Mizutani, 2009; Spratt, 2005; Wall, 2000).

Il est alors nécessaire d'identifier et de comprendre les facteurs qui influencent ces effets. Les effets de l'évaluation externe fonctionnent par un « mécanisme » complexe incluant tous les niveaux de contexte et les acteurs de la vie éducative. Les caractéristiques des épreuves doivent donc être prises en compte, mais il ne faut pas oublier celles du contexte et des acteurs (Alderson, 2004; Shohamy, Donitsa-Schmidt, & Ferman, 1996; Wall, 2000). Ces études soulignent notamment les différences entre les enseignants et l'importance de leurs croyances (*teachers' beliefs*). Des modèles qui identifient ces facteurs sont proposés dans la littérature. Nous en présentons quelques-uns que nous considérons comme les plus complets et nous permettent par la suite d'en faire une synthèse :

- Qi (2005, 2007) propose une analyse des effets d'un test standardisé national d'anglais sur les pratiques des enseignants en Chine. Son étude identifie les facteurs suivants : 1) la forme du test (par exemple : multiple-choice) ; 2) les finalités du test ; 3) la forme de communication des résultats à la société ; 4) les enjeux du test ; 5) les croyances (*beliefs*) des enseignants ; 6) l'expérience professionnelle des enseignants.
- Dans sa métaanalyse de la littérature consacrée au *washback*, Watanabe (2004) propose cinq groupes de facteurs : 1) les facteurs liés au test (contenu, forme) ; 2) les facteurs de prestige

(enjeux et statut du test) ; 3) les facteurs personnels (formation des enseignants et leurs croyances par rapport à l'enseignement et à l'apprentissage) ; 4) les facteurs micro (caractéristiques des écoles) ; 5) les facteurs macro (contexte du pays, région).

- Dans sa revue de littérature sur le *washback*, Spratt (2005) distingue cinq catégories de facteurs: 1) les facteurs liés aux enseignants ; 2) les facteurs liés aux ressources ; 3) facteurs liés à l'école ; 4) facteurs liés au test. Ces catégories sont détaillées dans la figure 7.

Teacher-related factors	Resource, the school, the exam
<p><i>Teacher beliefs about:</i></p> <ul style="list-style-type: none"> ● the reliability and fairness of the exam ● what constitute effective teaching methods ● how much the exam contravenes their current teaching practices ● the stakes and usefulness of the exam ● their teaching philosophy ● about the relationship between the exam and the textbook ● their students' beliefs <p><i>Teachers' attitudes towards:</i></p> <ul style="list-style-type: none"> ● the exam ● preparation of materials for exam classes ● lesson preparation for exam classes <p><i>Teachers' education and training:</i></p> <ul style="list-style-type: none"> ● Teachers' own education and educational experience ● the amount of general methodological training they have received ● training in teaching towards specific exams and in how to use exam-related textbooks ● access to and familiarity with exam support materials such as exam specifications ● understanding of the exam's rationale or philosophy. <p><i>Other:</i></p> <ul style="list-style-type: none"> ● personality ● willingness to innovate 	<p><i>Resources:</i></p> <ul style="list-style-type: none"> ● the availability of customised materials and exam support materials such as exam specifications ● the types of textbooks available <p><i>The school:</i></p> <ul style="list-style-type: none"> ● its atmosphere ● how much the administrators put pressure on teachers to achieve results ● the amount of time and number of students allocated to exam classes ● cultural factors such as learning traditions <p><i>The exam:</i></p> <ul style="list-style-type: none"> ● its proximity ● its stakes ● the status of the language it tests ● its purpose ● the formats it employs ● the weighting of individual papers ● when the exam was introduced ● how familiar the exam is to teachers

FIGURE 7 – Modèle de *Washback* de Spratt (2005, p.29)

- De son étude qualitative (par observations et interviews) sur les effets des épreuves standardisées d'anglais dans des écoles de Taiwan, Shih (2009) identifie différents groupes de facteurs qui ont un effet sur le « degré » de *washback* sur l'enseignement : 1) facteurs liés au contexte ; 2) facteurs liés au test ; 3) facteurs liés aux enseignant. Ces trois facteurs sont détaillés dans la figure 8.

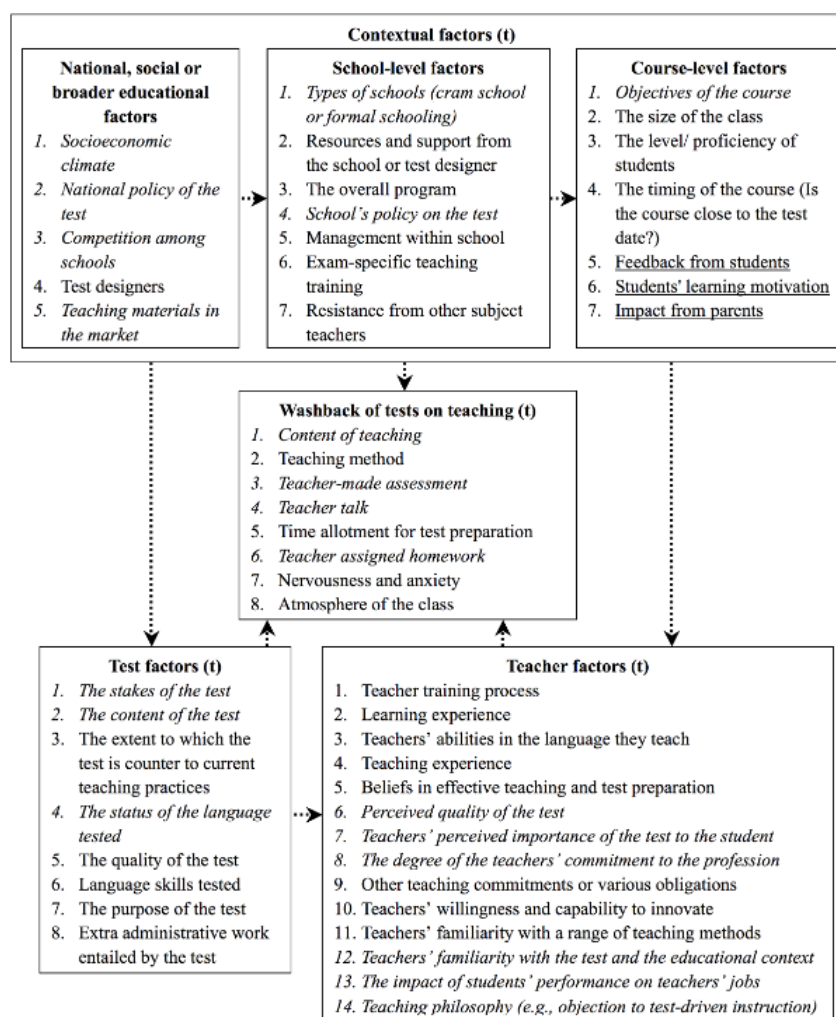


FIGURE 8 – Modèle du Washback effect de Shih (2009, p.199)

Les facteurs en italique sont identifiés à la fois dans son étude et par d'autres études. Les facteurs soulignés sont uniquement identifiés dans sa recherche ; les autres sont identifiés par d'autres recherches. Outre cette catégorisation, ce modèle donne la direction de ces influences. L'effet des tests est influencé par les trois catégories de facteurs. De plus, des effets médiateurs sont observés, les facteurs de contexte influençant également ceux du test et des enseignants. Le symbole (t) démontre que le temps a aussi un impact sur tous ces facteurs.

- Mizutani (2009) propose une analyse quantitative des effets d'un test standardisé national sur les pratiques des enseignants en Nouvelle-Zélande. Elle donne une nouvelle perspective. Au contraire des recherches qualitatives citées précédemment, elle permet d'observer directement leur importance et le degré de leurs interactions. Les résultats de Mizutani (2009) montrent que certains facteurs liés au contexte et aux enseignants influencent les effets de l'évaluation externe sur leurs pratiques (figure 9).

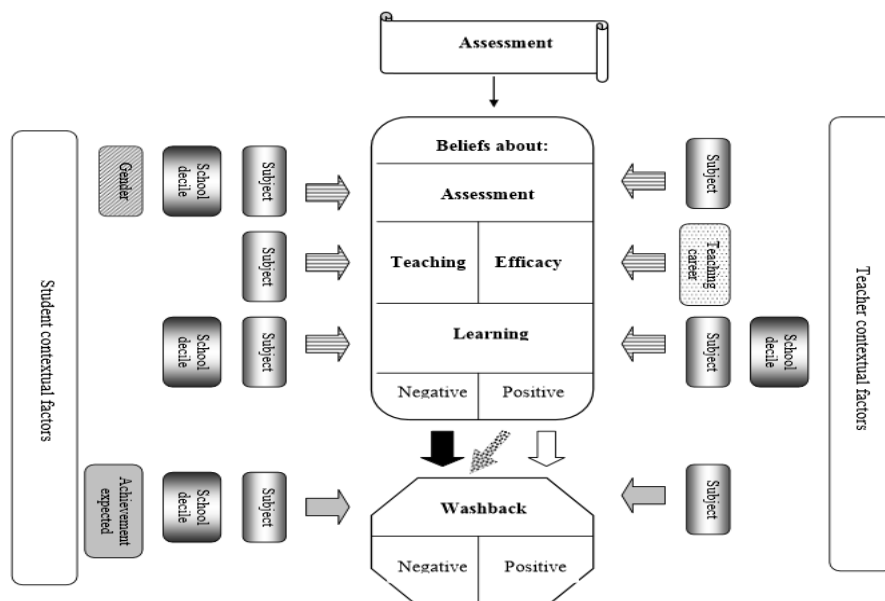


FIGURE 9 – Modèle proposé par Mizutani (2009, p.207)

Ces facteurs peuvent être en interactions et avoir un impact direct (flèche unie) ou indirect (flèche bicolore) sur l'effet du test (*washback*) :

- *Effets directs.* Les *beliefs* des enseignants influencent directement le *washback*. De plus, si les *beliefs* sont positifs, ils entraînent des effets positifs et négatifs ; si les *beliefs* sont négatifs, ils entraînent des effets négatifs. La branche d'enseignement est aussi un facteur d'influence direct, les enseignants de branches scientifiques étant davantage touchés que ceux enseignants des branches littéraires.
- *Effets indirects.* La branche enseignée peut avoir un effet sur les conceptions d'évaluation et de l'apprentissage. L'expérience professionnelle sur l'efficacité. Le statut socioéconomique de l'école a un impact sur les conceptions de l'apprentissage des enseignants.

En nous inspirant de ces modèles, nous détaillons ci-dessous trois catégories de facteurs qui sont présents dans la plupart de ces modèles : 1) les facteurs liés au contexte ; 2) les facteurs liés au test ; 3) les facteurs liés aux enseignants. Ces facteurs peuvent d'ailleurs être interdépendants comme le montre les modèles de Shih (2007) et de Mizutani (2009).

7.1. Les facteurs liés au contexte

Selon les modèles présentés mais aussi dans la littérature plus générale, le contexte politique et éducatif dans lequel s'inscrivent les épreuves externes est cité comme un facteur d'influence sur les effets des évaluations externes sur les pratiques des enseignants. En effet, les dispositifs de *testing* et les usages des épreuves externes sont différents dans les pays ou régions (voir notamment Eurydice, 2009; Looney, 2009; Soussi & Nidegger, 2012). Leurs effets dépendent donc des différentes « configurations institutionnelles » et des divers modes de régulation de chaque contexte (Maroy, 2013; Mons, 2009). Ces différences sont observées à divers niveaux:

- *Niveau macro*. Les effets de l'évaluation externe varient selon les décisions politiques liées à divers éléments tels que leur temporalité et leur fréquence dans la scolarité, leurs fonctions pédagogiques et/ou politiques, les usages et la publication de leurs résultats, leur inscription dans des politiques d'*accountability*, ...
- *Niveau méso*. Les effets de l'évaluation externe varient selon les décisions et le management des cadres intermédiaires (responsables éducatifs locaux, inspecteurs, directeurs d'école) liés aux résultats des épreuves externes.
- *Niveau micro*. Les effets de l'évaluation externe varient selon les caractéristiques de l'établissement (type d'école, ressources, climat de travail entre enseignants...) ou de la classe (effectifs, niveau, etc.).

L'étude de Dumay et al. (2013)³⁵ montre que les modes d'organisation des écoles (ici en mode « réseau » ou « bureaucratie ») peuvent avoir une influence sur les effets de l'évaluation externe. Une même politique (au niveau macro) peut effectivement être mise en place de manières différentes par les cadres intermédiaires (au niveau méso) et avoir, au final, des résultats différents sur le terrain (au niveau micro). Les auteurs concluent en relevant l'importance de l'échange entre ces différents niveaux contextuels³⁶.

7.2. Les facteurs liés à l'épreuve externe

Les spécificités des tests eux-mêmes ont une influence sur la variabilité des effets en classe. Ces facteurs dépendent évidemment du contexte politique et éducatif dans lequel s'inscrit le dispositif. Dans les modèles présentés, plusieurs caractéristiques de l'épreuve sont citées et peuvent se trouver en interaction :

³⁵ Cette analyse reprend les données de l'étude de Cattonar et al. (2010).

³⁶ Voir également les conditions relevées par Coburn (2004) et Laveault (2009b) (p.60-61).

- *Ses enjeux.* En lien avec les politiques dans lesquelles elle s'inscrit, l'épreuve peut avoir des enjeux de divers degrés pour les enseignants ou les élèves. De nombreux auteurs et de nombreuses recherches notent cette différence, les dispositifs entraînant des enjeux forts sont les plus prompts à avoir des effets (positifs et/ou négatifs) sur les acteurs de la scolarité (Crahay & Lafontaine, 2012). Pourtant d'autres recherches ne trouvent pas de différences. Les résultats d'une analyse de 25 Etats américains (dont les politiques d'accountability sont analysées et classées de *low* à *high*) montrent que les enjeux n'ont pas d'influence sur les performances des élèves (Nichols et al., 2006).
- *Sa forme.* La forme de l'épreuve peut influencer les enseignants à s'en inspirer ou non. Elle est notamment déterminante en ce qui concerne le type de questions et de tâches demandées aux élèves (question ouvertes ou fermée, niveaux taxonomiques,...).
- *Son contenu.* L'épreuve externe peut avoir des effets différents selon son contenu. Il peut être « aligné » ou non sur les programme d'études. La quantité de contenu peut aussi être variable (une ou plusieurs branches, contenu d'une année, d'un cycle, d'une période...).
- *Ses finalités.* Les finalités de l'évaluation peuvent être multiples (comme présentées dans le point 5, p.33-47). Il a notamment été démontré que les différentes finalités peuvent parfois se court-circuiter. L'étude de Qi (2005) démontre que les intentions de régulation des pratique d'une épreuve externe échouent surtout à cause de sa double fonction de sélection et de régulation. La fonction de sélection empêche la régulation des pratiques : elle empêche les concepteurs de montrer une évaluation de qualité, aussi les enseignants occultent la fonction de régulation ne voyant uniquement la fonction de sélection.
- *Sa qualité.* La qualité de l'épreuve dépend de son contenu et sa forme.
- *Sa temporalité.* L'épreuve peut intervenir dans différents degrés de la scolarité, mais aussi dans diverses périodes d'une année scolaire. Cette temporalité peut entraîner des différences d'effets dans les classes.
- *La charge de travail imputable à l'épreuve.* L'épreuve peut représenter une charge de travail variable pour l'enseignant (passation, correction, préparation, ...).

7.3. Les facteurs liés aux enseignants

Bien que les politiques aient certaines attentes quant aux effets des épreuves externes, l'« institutionnalisation » de leur message dépend surtout de la réception qui en est faite sur le terrain (Dumay et al., 2013). Les modèles présentés précédemment relèvent des facteurs qui ont trait précisément au profil de l'enseignant :

- *Profil personnel et professionnel.* Parmi les facteurs liés aux enseignants, rares sont ceux qui traitent directement de leur profil personnel (sexe, âge, ...). C'est surtout l'expérience professionnelle qui est citée comme pouvant influencer les effets de l'évaluation externe. Elle peut être de plusieurs ordres : l'expérience d'enseignement et l'expérience d'épreuves externes. D'autres aspects sont également identifiés comme celui de leur réaction face à l'innovation et la branche d'enseignement.
- *Les croyances des enseignants (teachers' beliefs).* Les modèles exposés précédemment relèvent que les croyances des enseignants ont une influence importante et directe sur les effets de l'évaluation externe. La majorité des études qui en traitent sont anglophones. Ce terme s'oppose aux connaissances (*knowledges*) et peut être traduit par de nombreux termes francophones qui sont généralement considérés comme interchangeables : représentations, théories personnelles, conceptions, théories implicites... (Crahay, Wanlin, Issaeva, & Laduron, 2011). Nous employons donc ces différents termes comme synonymes qui désignent « une représentation que se fait un individu de la réalité ; celle-ci possède assez de validité, de vérité, ou de crédibilité pour guider la pensée et le comportement » (Harvey, 1986, cité par Pajares, 1992, p. 313). Ces croyances résultent des expériences personnelles et professionnelles des enseignants et leur servent de références pour l'analyse et la gestion de situation d'enseignement. Elles auraient donc un impact direct sur les pratiques des enseignants, étant un filtre entre les exigences officielles et les pratiques effectives. La littérature différencie généralement des types de croyances (Crahay et al., 2011). Justement, dans les modèles présentés, plusieurs types de croyances des enseignants sont relevés comme ayant une influence sur les effets des épreuves standardisées. Celles qui sont le plus citées sont d'ordre « pédagogique générale » : leurs conceptions de l'efficacité pédagogique, leurs conceptions de l'apprentissage, leurs conceptions de l'enseignement, leurs conceptions de l'évaluation scolaire.
- *Perceptions de l'épreuve externe par les enseignants*³⁷. Dans certains modèles, les perceptions de l'évaluation externe par les enseignants font directement partie de leur système de croyances, dans d'autres elles relèvent de leurs attitudes. La façon dont les enseignants perçoivent le dispositif d'évaluation externe influence les effets de celui-ci sur leurs pratiques. Cette perception est en lien avec les facteurs propres à l'épreuve : ses enjeux pour les élèves et les enseignants, sa qualité, son utilité, la charge de travail... Pourtant les enseignants perçoivent différemment ces facteurs, ils deviennent alors de facteurs perçus (enjeux perçus, qualité perçue, utilité perçue...). Les recherches dans le champ du *washback* démontrent

³⁷ Les perceptions de l'évaluation externe par les enseignants sont l'objet du point 8 (p.83-84).

d'ailleurs que ce ne sont pas les épreuves elles-mêmes qui influencent les pratiques des enseignants, mais plutôt leurs perceptions des changements introduits par ces dispositifs. Le changement n'est pas « mécanique », il dépend du message émis par les autorités (lié au contexte), mais surtout de sa réception par les enseignants au travers de leurs conceptions (Chapman & Snyder, 2000; Mizutani, 2009).

8. Perceptions de l'évaluation externe par les enseignants

Comme démontré au point précédent par les modèles du *washback effect*, les perceptions de l'évaluation externe par les enseignants peuvent prendre une place importante dans l'explication de leurs effets. La littérature sur le regard porté par les enseignants sur ces dispositifs est riche et ce dans différents contextes. Elle montre cependant une certaine ambiguïté. Les perceptions des enseignants quant aux épreuves externes sont généralement contrastées et ce dans tous les contextes étudiés. Il s'agit souvent d'un "mal nécessaire". Les enseignants sont à la fois favorables à ces outils, mais craignent leur utilisation (Mons, 2009).

8.1. Les enseignants plutôt favorables

Des sondages dans différents pays ou régions (USA, Scandinavie, Suisse romande, France, Belgique) montrent que les enseignants apprécient pour la plupart les évaluations standardisées pour le rôle de modèle d'évaluation (mesure objective et équitable, comparaisons) et de guide (clarification des contenus et des niveaux de performance attendus) (Mons, 2009). Premièrement, comparée à l'évaluation faite par l'enseignant dans sa classe (interne), l'évaluation externe est souvent considérée comme une mesure davantage « objective, rigoureuse et digne de confiance », même par les enseignants (Lafontaine et al., 2009). Des recherches, menées en Suisse romande notamment, ont démontré que les enseignants sont majoritairement favorables à de tels dispositifs afin de vérifier et de compléter leur évaluation interne (Dierendonck, 2008; Longchamp, 2010; Ntamakiliro & Tessaro, 2002; Soussi et al., 2006). L'évaluation externe permet aux enseignants de situer les performances de leurs élèves (et de ce fait leur propre performance) en comparaison à une norme (Lafontaine et al., 2009; Tessaro & Ntamakiliro, 2010a). Elle donne alors une référence pour « situer leur appréciations et leurs degrés d'exigence ou de sévérité » (Weiss, 1994, cité par Mottier Lopez & Laveault, 2008, p. 21). Deuxièmement, outre le fait de pouvoir situer les performances de leurs élèves, les plupart des enseignants sont également favorables aux épreuves externes qui leurs servent de « lignes claires » pour bien comprendre les contenus qui sont attendus à la fin de chaque degré ou cycle (Mons, 2009). Finalement, les épreuves externes peuvent être vues comme « rassurantes », permettant aux enseignants valider leurs pratiques (Cattonar et al., 2010).

Au total, la perception générale par les enseignants des standards et des dispositifs de tests associés ainsi que dans, certaines configurations institutionnelles, la réalité des effets de ces mesures sur l'activité pédagogique semblent positives : ces réformes permettent en effet de donner des guides clairs pour mettre en œuvre les curricula, d'empêcher l'apparition de fortes inégalités dans le développement des syllabi locaux, de mettre l'accent sur les résultats réels des élèves, en particulier pour ceux issus des milieux défavorisés, et de favoriser un travail en équipe autour de l'analyse des résultats des évaluations (Mons, 2009, p.121).

8.2. Les résistances des enseignants

Si l'évaluation externe en tant qu'outil pédagogique est plutôt bien acceptée par les enseignants, des sondages montrent que les modalités de son utilisation font davantage l'objet de résistances (Mons, 2009). Premièrement, dans plusieurs pays, la majorité des enseignants se plaint du nombre d'épreuves externes trop important (par ex. USA, Angleterre, France). Deuxièmement, les enseignants peuvent remettre en cause les résultats des évaluations externes, leur rapprochant leur caractère trop standard et quantitatif. Enfin, l'évaluation externe est surtout rejetée par les enseignants lorsque l'utilisation de ses résultats comprend des enjeux trop élevés, pour les élèves (promotion, sélection) ou pour eux-mêmes (contrôle, récompenses, sanctions). D'ailleurs, si les enseignants s'opposent à ce genre de dispositifs c'est généralement qu'ils se sentent contrôlés, ayant le sentiment que leurs compétences et leur autonomie étant mises en doute (Dierendonck, 2008; Helgoy & Homme, 2007; B. Jones, 2007; Lafontaine et al., 2009; Pons, 2011; Soussi et al., 2006).

Des études, surtout dans la littérature anglo-saxonne, démontrent ce dernier argument. Elles témoignent surtout d'un stress important chez les enseignants confrontés aux évaluations externes (Mons, 2009). L'étude de Cattonar et al. (2010) montre que les pressions ressenties par les enseignants sont variables. Elles diffèrent selon le degré d'enseignement, le type d'épreuves (certificatives ou non) ou l'expérience professionnelle (les premières expériences étant davantage stressantes). Ce sentiment de stress est tout de même lié au sentiment d'être évalué soi-même et aux enjeux des résultats, tels que la réputation personnelle ou celle de l'école, la réussite des élèves, les attentes des parents, la pression des collègues ou de la direction. Ce stress s'explique également par la pression due aux échéances. Les enseignants doivent réussir à traiter toute la matière évaluée. Cette recherche montre par contre que peu d'enseignants voient leur autonomie restreinte par les épreuves, le discours des enseignants étant quelque peu « ambivalent » : ils considèrent bénéficier d'une grande autonomie, mais les épreuves les contraignent à enseigner d'une certaine façon (ici les compétences). Dans le canton de Genève, l'étude de Soussi et al. (2006) montre que les enseignants des degrés primaires et secondaires ne se sentent en moyenne pas évalués eux-mêmes par les épreuves cantonales externes. En CFB, au contraire, la majorité des enseignants du degré primaire interrogés par questionnaire (65%) se sentent « contrôlés » par l'évaluation externe, mais que leur autonomie n'est pas touchée (80%). De même ils sont peu nombreux à penser que l'évaluation externe « les décourage » (moins de 10%). L'évaluation externe ne représente pas non plus pour eux une charge de travail trop importante (40% des enseignants de 5^{ème} primaire ; 27% des enseignants de 2^{ème} primaire) (Lafontaine et al., 2009). Des recherches du champ du *washback* montrent également que les enseignants ressentent de fortes pressions face aux épreuves externes. Ce sentiment de stress dépend des enseignants et des caractéristiques des tests (Spratt, 2005).

9. Recommandations pour la mise en place de l'évaluation externe

Même si aucun modèle « idéal » d'évaluation externe n'existe (Mons, 2009), nous retenons ici certaines recommandations proposées par certains chercheurs pour la mise en place d'épreuves externes, afin d'éviter leurs effets néfastes.

a) *Une évaluation davantage « démocratique »*

L'intégration d'une épreuve externe dans un système éducatif est souvent justifiée comme « une solution technique de bon sens » (Mons & Pons, 2006). Pourtant des enjeux importants en découlent et elle devrait donc être le fruit d'une réflexion systémique à plusieurs niveaux (micro, méso, macro) quant à ses fonctions et à ses éventuels effets sur tous les acteurs. Pourtant, il semble que c'est rarement le cas. L'évaluation externe fait le plus souvent partie de l'« évaluation gestionnaire », propre à l'administration et qui n'inclut pas les autres acteurs dans les prises de décisions, au contraire d'une « évaluation démocratique » (Duran & Monnier, 1992). Dans ce sens, il est aussi important de déterminer les responsabilités des chaque acteur, quant aux résultats de l'épreuve (Mons & Pons, 2006).

b) *Usage des résultats de l'épreuve*

Les effets de l'évaluation externe sur tous les acteurs dépendent surtout de l'usage qui est fait de leurs résultats. C'est le mode de régulation (ou *accountability*) qui définit ses modalités, notamment par la manière de les communiquer : soit les résultats sont utilisés de manière externe et interne, soit ils restent internes uniquement (Mons & Pons, 2006) :

- *Externe et interne.* Les résultats de l'épreuve externe sont communiqués de manière simplifiés à la société (notamment aux parents), souvent sous forme de classement ou de palmarès.
- *Interne.* Les résultats sont surtout communiqués aux autorités, aux établissements et aux enseignants de manière détaillée ou non. La façon de communiquer ces résultats reste anonyme, les établissements et les enseignants pouvant se comparer aux autres sans toutefois être nommés pour éviter toute forme de compétition. Dans une moindre mesure les résultats sont communiqués aux parents (sous la forme de note, mais sans comparaison directe à la norme).

c) *Intégration de l'épreuve dans un « maillage » d'évaluations*

Comme exposé auparavant, pour de nombreux chercheurs (certains économistes), l'évaluation de la qualité d'un système éducatif (et de toute politique publique) ne peut se faire uniquement avec des résultats quantitatifs. Il en va de même pour l'évaluation externe des acquis des

élèves. Ces épreuves devraient être accompagnées de mesures qualitatives complémentaires (Paul, 2005). Pour Laveault (2009a), plusieurs indicateurs doivent être pris en compte, pas uniquement les résultats mais aussi les inputs, les processus et le contexte.

De plus, dans de nombreux systèmes éducatifs, les évaluations externes se multiplient à tous les niveaux (international, national, local) et leur nombre devient trop élevé et leurs rôles sont quelque peu incompréhensibles. Ceci peut aussi expliquer leur rejet par le terrain.

Toutefois, l'analyse des pratiques montre que l'équilibre entre les deux finalités est difficile à réaliser. À différents niveaux du processus d'évaluation, la logique de l'évaluation des élèves d'un côté, celle du pilotage et du monitoring du système d'enseignement de l'autre, s'opposent et la fiabilité des épreuves externes en est affectée (Ntamakiliro & Longchamp, 2008, p. 1).

d) *Alignement des contenus*

Les épreuves doivent être construites en lien avec les objectifs des plans d'études et les standards. Leur contenu doit être fidèle aux référentiels, c'est ce qu'on appelle l'« alignement ». De plus, les résultats doivent être suffisamment détaillés selon les composantes d'une matière spécifique ou d'une habileté (détailler par exemple en mathématiques le niveau de géométrie) (Laveault, 2009a). Toutefois, il est observé que les épreuves évaluent pour la plupart qu'une partie des programmes, certaines branches « prioritaires » étant privilégiées (langues et mathématiques). De plus, on constate que les tests externes se centrent de plus en plus sur des compétences et non sur des contenus disciplinaires (Eurydice, 2009).

e) *Conception rigoureuse et validité scientifique*

Les tests externes sont conçus soit par les autorités elles-mêmes, soit par des organes liés aux autorités (incluant des enseignants et spécialistes), soit par des organes privés (Eurydice, 2009). Wirthner & Ntamakiliro (2008) observent que les épreuves conçues uniquement par des spécialistes sont moins bien reçues par les enseignants, mais gagnent en qualité surtout pour ce qui est de la standardisation. La solution serait donc dans la collaboration entre enseignants et experts. Aussi, l'épreuve devrait être validée par un pré-test, d'une passation d'essai par un échantillon représentatif d'élèves (Ntamakiliro, 2003). Toutefois, l'élaboration et l'implantation d'épreuves externes sont bien peu souvent légitimées par des travaux scientifiques (Mons & Pons, 2006). Aussi, certains chercheurs soulignent la nécessité que les évaluations externes soient davantage « longitudinales », afin de pouvoir observer le progrès des systèmes et des élèves (en prenant plusieurs mesures sur une période définie). Cette option

va de pair avec celle de « valeur ajoutée » : il faut pouvoir estimer l'amélioration des élèves (Laveault, 2009a; Paul, 2005).

f) *Standardisation des conditions de passation et de correction des tests*

Pour être équitable et pouvoir donner des informations pertinentes et comparables, l'épreuve doit être « standardisée ». Elle doit suivre les règles³⁸. Généralement, les épreuves externes sont administrées aux élèves par les enseignants, à qui on met à disposition des consignes de passation standards. Mais dans certains cas, une personne externe à la classe (souvent un autre enseignant) en est responsable. Il en va de même pour la correction.

g) *Traitement rigoureux des résultats*

Tous les résultats des élèves, ou en tout cas d'un échantillon, doivent être enregistrés dans une base de données et une analyse en est faite (Ntamakiliro, 2003). Le traitement des informations doit se faire par une analyse scientifique et des statistiques appropriées (Laveault, 2009a).

h) *Impact budgétaire*

Le coût des évaluations standardisées est souvent occulté par les autorités. Toutefois, ce principe est contraire à leurs propres origines : ces dispositifs d'évaluation sont créés au nom de l'efficacité des systèmes et à la « reddition de comptes » aux parents et à la société toute entière. Les coûts de telles évaluations devraient alors être eux aussi mis en perspectives de leurs apports (Mons & Pons, 2006).

³⁸ Les conditions pour la standardisation d'un test sont présentées au point 2.2. (p.7-8).

10. Questions de recherche et hypothèses

Le cadre théorique donne, par la présentation de travaux de chercheurs issus de contextes et de champs divers, de premières réponses aux questions énoncées dans la problématique de la thèse. Ces différentes sources permettent d’affiner notre questionnement et d’identifier certains manques. La nécessité de données empiriques est notamment soulevée en ce qui concerne les effets de l’évaluation externe des acquis des élèves sur les pratiques des enseignants, surtout afin d’en comprendre les processus. Ce constat nous permet ainsi de formuler les questions de recherche de la thèse (QR). Avant de les présenter, nous les situons sur la figure 10, inspirée des modèles théoriques présentés dans le chapitre 1 (p.76-78). Elle représente les variables dépendantes (trait hachuré) et indépendantes (trait continu) que nous investiguons. Les flèches représentent les liens entre les variables³⁹ que nous allons interroger. Les facteurs de contexte et leurs liens avec les autres variables (pointillés) ne sont pas pris en compte dans cette recherche⁴⁰.

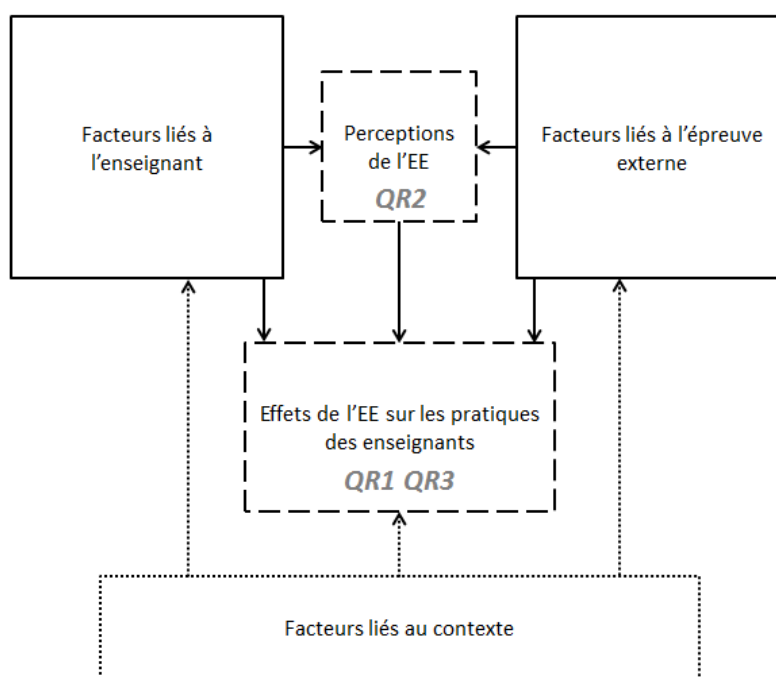


FIGURE 10 – Modélisation a priori des effets de l’évaluation externe (EE) et des facteurs qui les influencent

³⁹ Les flèches montrent la direction des liens, elles signifient “fait varier”.

⁴⁰ L’absence de facteurs liés au contexte dans notre étude est relevée dans les limites de notre travail (p.234)

Afin de répondre aux questions de recherche, nous formulons des hypothèses (H) qui leur sont sous-jacentes. Il s'agit donc de propositions de réponse à vérifier directement sur le terrain (Albarello, 2003).

QR1

Comment les enseignants du primaire perçoivent-ils l'évaluation externe des acquis des élèves ?

H1.1

Les enseignants sont favorables à l'évaluation externe des acquis des élèves mais craignent l'usage de ses résultats.

H1.2

Les perceptions des enseignants quant à l'évaluation externe des acquis des élèves dépendent des finalités de l'épreuve externe.

H1.3

Les perceptions des enseignants quant à l'évaluation externe des acquis des élèves dépendent de leur expérience professionnelle.

H1.4

Les perceptions des enseignants quant à l'évaluation externe des acquis des élèves dépendent de leurs croyances à propos de l'enseignement.

QR2

Quels sont les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants du primaire ?

H2.1

L'évaluation externe des acquis des élèves affecte les pratiques des enseignants dans tout le processus d'enseignement (planification de l'enseignement, enseignement-apprentissage, évaluation de l'apprentissage).

H2.2 L'évaluation externe des acquis des élèves produit des effets plus grands sur les pratiques des enseignants lorsqu'elle est accompagnée de forts enjeux.

H2.3 Pour un même type d'évaluation externe des acquis des élèves, les enseignants ne s'accordent pas sur l'importance de ses effets sur leurs pratiques.

H2.4 Des liens existent entre certains effets de l'évaluation externe des acquis des élèves.

QR3

Quels sont les facteurs qui influencent les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants du primaire ?

H3.1 Les facteurs liés à l'épreuve externe influencent les effets de l'évaluation externe sur les pratiques des enseignants.

H3.2 Les facteurs liés aux perceptions des enseignants quant à l'épreuve influencent les effets de l'évaluation externe sur leurs pratiques.

H3.3 Les facteurs liés à l'enseignant influencent les effets de l'évaluation externe sur leurs pratiques (expérience professionnelle et croyances à propos de l'enseignement).

11. Articulation de notre recherche et choix méthodologiques

11.1. Choix de l'approche méthodologique de notre recherche

L'objectif de la thèse est de répondre à nos questions de recherche en vérifiant les hypothèses sous-jacentes sur le terrain⁴¹. Pour ce faire, des méthodes diverses et complémentaires sont nécessaires. Alors que la question de l'évaluation externe fait appel à différentes disciplines (notamment la sociologie, l'évaluation des programmes, la politique et l'économie), nous avons fait le choix d'une approche principalement psychologique. Ce choix est déterminé par les variables investiguées, notamment les perceptions des participants. Nos hypothèses suggèrent différentes « techniques de recherche » (Albarello, 2003), les variables dépendantes ou indépendantes, et leur interdépendance, étant d'une part identifiées (approche qualitative) et d'autre part mesurées (approche quantitative). Afin de répondre à nos questions, la thèse comprend deux parties empiriques distinctes dont les approches sont de natures différentes mais complémentaires⁴² :

- a) Dans un premier temps, *la première partie empirique exploratoire* (approche qualitative) permet d'identifier, dans le contexte de la recherche, les variables dépendantes et indépendantes et leur interdépendance (chapitre 1) ;
- b) Dans un deuxième temps, *la deuxième partie empirique* (approche quantitative) permet de vérifier et de généraliser, dans le contexte de la recherche, la présence de ces variables et leur interdépendance sur un échantillon plus large. Elle permet d'observer leur variabilité, d'en identifier les différences selon les différents facteurs et de mesurer le rôle des facteurs explicatifs (chapitre 2) ;
- c) Dans un troisième temps, comme elles répondent aux mêmes questions de recherche, les données des deux parties (qualitatives et quantitatives) sont analysées (chapitre 5) et discutées (chapitre 6) conjointement pour vérifier les hypothèses et formuler des réponses aux questions de recherche.

La méthode globale utilisée dans la thèse peut être décrite comme « méthode mixte ou *mixed method research* » (Johnson & Onwuegbuzie, 2004; Pluye et al., 2012).

En général, les méthodes mixtes sont utilisées en vue de combiner les forces respectives des méthodes quantitatives et qualitatives. D'une part les évaluations quantitatives examinent habituellement des associations de facteurs (par exemple, des causes et leurs effets) pouvant être généralisées à toute une population, notamment grâce à des inférences statistiques. [...] D'autre part, les évaluations qualitatives

⁴¹ Le contexte de la recherche est décrit dans le chapitre 2 (p.93-110).

⁴² La méthodologie et les résultats de chaque partie empirique sont présentés séparément dans les chapitres 3 et 4.

fournissent des descriptions détaillées de phénomènes complexes fondées empiriquement sur un contexte spécifique pouvant être théoriquement applicables dans d'autres situations (Pluye et al., 2012, p. 125).

Il ne s'agit en effet pas de deux études distinctes, mais bien d'une complémentarité entre ces deux approches, l'une permettant de comprendre (qualitatif) et l'autre de mesurer (quantitatif). Ainsi, les différentes approches servent l'une à l'autre.

- a) les résultats quantitatifs permettent de généraliser les résultats qualitatifs ;
- b) les résultats qualitatifs permettent d'interpréter les résultats quantitatifs.

11.2. Choix du contexte de notre recherche

Pour mener à bien nos objectifs et tenter de répondre à nos questions de recherche, nous avons choisi de mener notre étude sur un contexte scolaire particulier, celui de la scolarité primaire du canton de Fribourg (Suisse)⁴³. Ce choix s'explique tout d'abord car ce système scolaire comprend deux types distincts d'épreuves externes des acquis des élèves. Ensuite, il est déterminé par notre expérience et notre très bonne connaissance du contexte fribourgeois et par notre aptitude à créer un lien de confiance avec les enseignants⁴⁴.

12. Implications théoriques et pratiques de la thèse

Sur le plan théorique, cette thèse de doctorat pourrait permettre de :

- compléter les connaissances théoriques à propos des effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants ;
- combler un manque dans la recherche : comprendre le processus à l'origine de ces effets;
- créer des outils pour mesurer ces effets et qui puissent être utilisés dans différents contextes.

Les résultats de la thèse pourraient être utilisés dans la pratique, pour :

- informer les concepteurs et administrateurs d'épreuves externes des possibles effets des évaluations externes standardisées sur les pratiques des enseignants;
- adapter la conception et l'utilisation d'épreuves externes ;
- accompagner les enseignants dans la compréhension et l'utilisation des épreuves externes.

⁴³ Le contexte et les épreuves externes qui y sont inscrites sont décrits dans le chapitre 2 (p.93-110).

⁴⁴ L'auteur, en parallèle de ses activités de recherche et d'enseignement universitaires, est enseignant au degré primaire dans le canton de Fribourg.

CHAPITRE 2

CONTEXTE DE LA RECHERCHE

13. Développement de l'évaluation externe en Suisse.....	95
13.1. Au niveau fédéral	95
13.2. Au niveau régional	98
13.3. Au niveau cantonal	99
14. L'évaluation externe dans le canton de Fribourg.....	103
14.1. Organisation de la scolarité primaire fribourgeoise.....	103
14.2. Les différentes épreuves externes du canton de Fribourg.....	104
14.2.1. Epreuves cantonales de référence de fin de cycle	104
14.2.2. Epreuves cantonales de la procédure de pré-orientation (PPO)	105
14.2.3. Différences et classement des types d'épreuves cantonales fribourgeoises	107

13. Développement de l'évaluation externe en Suisse

Le système scolaire suisse a connu lors des dernières décennies de grands changements, au même titre que ceux d'autres pays de l'OCDE. L'une des mutations les plus importantes est l'avènement de l'évaluation externe des acquis des élèves comme élément central de prise de décisions pour l'élève et l'enseignant (rôle pédagogique) et pour le système scolaire (rôle politique)⁴⁵. En Suisse, le développement de ces dispositifs s'est fait à tous les niveaux du système : au niveau fédéral, régional et cantonal. L'objectif de ce chapitre est de présenter cette évolution et de faire un état des lieux de la situation actuelle notamment dans le canton de Fribourg, contexte de notre recherche.

13.1. Au niveau fédéral

La Suisse est un état fédéral où les 26 cantons qui le composent sont souverains en matière d'éducation et de formation. Toutefois, lors des dernières décennies, de nombreuses réformes nationales ou régionales ont fait évoluer le paysage éducatif suisse vers une certaine uniformité, notamment en matière d'évaluation, sous l'influence d'organes nationaux ou régionaux (CDIP⁴⁶, CIIP⁴⁷, D-EDK⁴⁸). Ces mutations s'apparentent à celles des autres pays de l'OCDE et leur origine s'explique par les mêmes facteurs de société. Nous en retraçons les grandes lignes grâce aux travaux de Behrens (2010), de la CDIP (2004) et de Wirthner (2010).

Les questions scolaires sont restées longtemps une affaire uniquement cantonale. Cependant, dans les années 1970, apparaissent les premières traces d'une coordination scolaire au niveau suisse (concordat du 29 octobre 1970) qui règle surtout l'âge de la rentrée à l'école obligatoire (6 ans) et sa durée (9 ans). De plus larges coordinations se sont exercées entre cantons, surtout au niveau régional (regroupés par langue) autour d'aspect disciplinaires (maths, langues) mais aussi autour des dispositifs d'évaluation des apprentissages des élèves, principalement sur l'usage des notes. Ces coordinations furent poursuivies dans les années 1980. De nouvelles conceptions de l'évaluation (évaluation formative, différenciée de l'évaluation sommative) apparurent dans les débats, inspirées par des

⁴⁵ La situation d'autres pays quant à l'usage de l'évaluation externe est présentée dans le point 4 (p.27-32).

⁴⁶ La Conférence suisse des directeurs de l'instruction publique (CDIP) regroupe les 26 ministres cantonaux chargés de l'éducation.

⁴⁷ La Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP) regroupe les ministres chargés de l'éducation des cantons francophones et italophones.

⁴⁸ La Conférence suisse alémanique des directeurs de l'instruction publique (D-EDK) regroupe les ministres chargés de l'éducation des cantons alémaniques.

travaux internationaux (Scriven et Bloom notamment). Dès 1990, s'amorça une certaine harmonisation des systèmes scolaires des cantons romands. Elle concernait les durées d'études (six années d'école primaire) mais aussi les procédures de passage entre l'école primaire et le secondaire. Par cette harmonisation, une conception commune de l'évaluation fut introduite dans les textes et dans les pratiques surtout à l'école primaire. L'évaluation formative et critériée prit une place privilégiée, les certifications voyant leur importance et leur nombre diminuer. Les enseignants furent déchargés en partie des tâches d'orientation et de sélection des élèves. Ces nouvelles conceptions de l'évaluation furent aussi inscrites dans la formation des enseignants (initiale et continue). Toutefois, certains cantons romands mirent un frein à ces réformes, renonçant tout (Valais) ou en partie (Fribourg) à l'introduction de ces nouvelles pratiques.

Si l'évaluation externe existait déjà en Suisse (surtout pour la certification et la sélection des élèves), elle prit de nouvelles formes et de plus amples fonctions dans les années 2000 : épreuves de référence à l'intention des enseignants, épreuves de certifications, épreuves internationales (PISA, TIMSS...). Elles servent aujourd'hui au pilotage des systèmes éducatifs au niveau cantonal et prochainement au niveau régional et fédéral. Ces nouveaux instruments politiques sont issus d'une période d'instabilité de l'école et de son image par le grand public. En Suisse, comme dans les autres pays de l'OCDE, la qualité de l'école est alors remise en cause lors de cette période. C'est notamment l'étude PISA 2000 qui donna aux autorités suisses une certaine légitimité pour introduire des changements importants. En effet, dans cette étude, les élèves suisses se classent dans la moyenne des pays de l'OCDE en littérature, 25% de ceux-ci présentant des compétences faibles, voire insuffisantes. Les résultats montrent aussi des variations importantes entre cantons. Alarmés par ce constat (notamment par la presse), les autorités cherchèrent alors à introduire des instruments visant à améliorer l'efficacité des systèmes scolaires, tirés des modèles anglo-saxons du *New Public Management (NPM)*.

L'institution école, restée majoritairement étatique en Suisse romande, est donc directement confrontée à la redéfinition de l'action publique. Cette mise en question se traduit par l'introduction de nouveaux modes de gestion administrative qui s'inspirent de ce que l'on appelle communément la nouvelle gestion publique ou la gestion par mandat de prestation (Behrens, 2010, p. 118).

Les finances publiques doivent être gérées de manière efficace. La priorité n'est alors plus aux *inputs* (formation des enseignants, moyens d'enseignements...), mais aux *outputs* (résultats des élèves). Des changements importants vont alors avoir lieu dans les systèmes éducatifs au niveau national et cantonal, l'aspect bureaucratique disparaissant peu à peu :

- *Mutations dans le métier de l'enseignant.* Le métier de l'enseignant et sa formation connaissent des grands changements (accroissement des exigences et des compétences, davantage de collaboration, de formation continue, de réflexion sur ses pratiques, adaptation aux changements...).

- *Décentralisation.* Dans une perspective de décentralisation, l'établissement scolaire prend davantage d'importance. Une grande autonomie lui est offerte, des projets communs doivent s'y développer. Son organisation ressemble de plus en plus à une petite entreprise (comprenant une hiérarchie et l'évaluation de ses membres).

C'est dans cette perspective d'un nouveau mode de régulation de l'action publique et celle des études internationales qu'apparaît le projet suisse d'Harmonisation de la scolarité obligatoire (HarmoS) lancé par la CDIP (2004). Le nombre de cantons signataires requis étant atteint (10 cantons), le concordat HarmoS est entré en vigueur le 1er août 2009. Ce projet initie « un bouleversement profond du système suisse de formation » (Behrens, 2010, p.120). Il a pour but l'amélioration de la qualité de l'école obligatoire suisse, en élargissant le concordat de 1970 entre les cantons. Ce projet a un caractère contraignant obligeant les autorités cantonales à uniformiser les principales structures de leur système éducatif (si elles ne le font pas, la Confédération impose une solution). En outre, la Confédération se dote d'outils de contrôle de la qualité du système (standards de formation et évaluation standardisée).

La scolarité obligatoire relève intégralement des cantons. Avec HarmoS, ceux-ci prennent leurs responsabilités dans le cadre de la CDIP et se déclarent prêts à introduire une obligation de résultats dans un cadre méthodologique bien défini et dans un accord intercantonal contraignant (CDIP, 2004, p.5).

Le concordat HarmoS définit les structures principales du système :

- *Harmonisation des structures.* Le concordat vise l'harmonisation du début de la scolarité (4 ans), la durée des degrés scolaires (un degré primaire de 8 ans et un degré secondaire de 3 ans), l'aménagement de la journée d'école (imposition d'horaires blocs et de structures d'accueil), l'enseignement des langues nationales (introduction d'une deuxième langue nationale et de l'anglais à l'école primaire).
- *Harmonisation des contenus.* Le concordat vise l'harmonisation des contenus au niveau national en déterminant les compétences essentielles que les élèves doivent remplir à différents moments de la scolarité (2^{ème}, 6^{ème} et 9^{ème} année) pour quatre disciplines (langue locale, langues étrangères, mathématiques, sciences naturelles). Ces compétences à atteindre sont formulées sous forme de « standards nationaux de formation » (standards de performance, qui déterminent un niveau minimal de compétences ou de connaissances attendu pour un âge donné). HarmoS ne prévoit donc pas les moyens mais les résultats attendus, les décisions quant aux plans d'études et aux moyens d'enseignement sont laissées aux régions linguistiques.
- *Renforcement du pilotage.* Le concordat prévoit un pilotage national du système scolaire obligatoire sur la base des résultats des élèves. L'analyse de l'atteinte ou non des standards par

un échantillon national (recueil de diverses données qualitatives et quantitatives) sert à l'évaluation de l'efficacité du système éducatif. Cette évaluation servira à informer les autorités au sujet des problématiques prioritaires afin de « prendre des mesures de soutien ou les correctifs de niveau qui s'avèreront nécessaires » (CDIP, 2004, p.5).

Chaque canton qui ratifie le concordat HarmoS s'engage donc à modifier ses structures, à abandonner ses plans d'études au profit d'un référentiel régional et à contrôler que tous ses élèves atteignent les niveaux de performance prévus par les standards de formation.

Le concordat HarmoS, par ses structures (surtout les standards de formation), suscita de nombreuses réactions de la part des chercheurs suisses en Sciences de l'éducation⁴⁹.

En Suisse, au niveau fédéral, l'évaluation externe connaît donc une importante évolution. A l'image des autres contextes présentés dans notre travail⁵⁰ et aux éléments théoriques avancés⁵¹, en plus de sa fonction pédagogique, elle est devenue un outil de politique éducative central. Wirthner (2010) s'interroge toutefois sur cette double fonction⁵². L'évaluation scolaire serait aujourd'hui en tension (voire en conflit) entre une approche interne (confiée aux enseignants) et une approche externe (confiée aux politiques via leurs organes de recherches).

Ne court-on pas le risque d'une mesure déconnectée de ce qui se fait en classe – et dès lors peu utile pour la remédiation de l'enseignement ? Ou d'un enseignement inféodé aux instruments de mesure, consistant essentiellement à préparer les épreuves d'évaluation pour conduire les élèves à de bons résultats ? (Wirthner, 2010, p.29).

13.2. Au niveau régional

En Suisse, l'évaluation externe prend de l'importance au niveau national surtout dans un but de pilotage. Cette évolution se fait également remarquer au niveau régional, les cantons s'associant en région par le biais d'organes politiques trans-cantonaux (CIIP, D-EDK). Ceux-ci forment des « espaces de formation » et, en parallèle au concordat HarmoS, éditent ou sont en train d'éditer leurs plans d'études communs (plan d'études romand, Lehrplan 21, piano di studio), formulant les compétences à atteindre par les élèves à différents degrés de la scolarité. En parallèle, chaque région prévoit des moyens pour évaluer ces plans d'études, sous la forme d'évaluation externe.

⁴⁹ Nous dirigeons le lecteur vers la lecture de l'ouvrage collectif édité par Baeriswyl et Périsset Bagnoud (2008).

⁵⁰ La situation d'autres pays quant à l'usage de l'évaluation externe est présentée dans le point 4 (p.27-32).

⁵¹ Les diverses fonctions de l'évaluation externe sont l'objet du point 5 (p.33-48).

⁵² Ce constat est partagé par d'autres chercheurs dans des contextes divers. Il fait l'objet du point 5.4 (p.45-48).

Le plan d'études romand (PER) étant entré en vigueur, des épreuves externes communes sont en cours d'élaboration (épreuves romandes communes). Elles n'ont toutefois pas encore été introduites à ce jour. Ces épreuves externes sont inscrites dans la « Convention scolaire romande », en plus des tests de monitoring nationaux (selon les standards HarmoS). Ces dernières serviront à la vérification de l'atteinte des objectifs du plan d'études romand (PER) à la fin de chaque cycle (3 cycles). Ces épreuves étaient originellement prévues comme épreuves de références utiles pour fournir des « repères » aux enseignants (fonctions pédagogiques). Toutefois, leurs fonctions se dirigent aujourd'hui plutôt vers un outil de pilotage de type bilan. De plus, dans les textes, il est prévu que tous les élèves passent ces épreuves et non un échantillon (Behrens, 2010; Wirthner, 2010).

13.3. Au niveau cantonal

L'évaluation externe des acquis des élèves a gagné une grande importance au niveau cantonal également. Si le recours à des épreuves externes existe déjà depuis longtemps (notamment dans un but de promotion et/ou de sélection des élèves dans le secondaire inférieur), les épreuves standardisées se sont multipliées dans la plupart des cantons dans tous les degrés ou cycles de la scolarité. En effet, le concordat HarmoS (CDIP, 2004), parmi d'autres changements visant à harmoniser la scolarité obligatoire, instaure un pilotage national et cantonal basé sur les résultats du système (CSRE, 2014; Gather Thurler & Périsset, 2014). Outre leur rôle de certification des apprentissages des élèves, ces dispositifs sont aujourd'hui utilisés de manière courante pour le pilotage des systèmes scolaires cantonaux par une analyse de leurs résultats. Les épreuves cantonales servent notamment à harmoniser les pratiques des enseignants, surtout en matière d'évaluation (Ntamakiliro & Tessaro, 2010).

Chaque canton, souverain en matière d'éducation, gère en effet ses propres dispositifs. Les cantons romands ont fait le choix d'épreuves externes des acquis des élèves, communément appelées épreuves cantonales (EC). Les cantons suisses alémaniques ont plutôt privilégié le choix de l'évaluation externe de la qualité des établissements (comprenant de multiples indicateurs). Toutefois, les cantons alémaniques débattent de la possibilité d'intégrer les résultats des élèves à des épreuves standardisées. Certains cantons les ont d'ailleurs déjà introduits (Argovie, Zürich) (CSRE, 2014). Les figures 11 et 12 présentent un état des lieux dans les systèmes scolaires des cantons suisses romands. Dans la première, elles sont classées selon la typologie du Consortium romand⁵³ auquel nous avons ajouté un aspect chronologique qui montre leur développement important. Dans la deuxième, on peut voir les disciplines évaluées, les moments durant l'année scolaire et l'utilisation des résultats⁵⁴.

⁵³ Les différentes typologies de l'évaluation externe sont l'objet du point 5.3.2 (p.42-44).

⁵⁴ Les auteurs ont réalisé cette figure en interrogeant les responsables cantonaux des épreuves.

canton	année	1	2	3	4	5	6	7	8	9
Berne							0			
Fribourg			0		0		X	0	0	X
Genève			X		X		X	X	X	X
Jura					0		X		0	
Neuchâtel		0	0	0	0	0	X		0 ¹ X	
Valais			0 ¹	0 ¹	0 ¹ X	0 ¹	0 ¹ X		X	0 ¹ X ²
Vaud			X		X		X		X	X

X = épreuves à visée diagnostique
 0 = épreuves à visée sommative
¹ = passation libre
² = épreuves à visée certificative

nouvelles épreuves inscrites entre 2004 et 2013

FIGURE 11 – Epreuves cantonales en Suisse romande, année scolaire 2013-2014 (CSRE 2014, p.51)

Canton	Degré	Disciplines	Moment			Utilisation des résultats				
			1	2	3	1	2	3	4	5
Fribourg	2 ^e	Ma			▲					▲
	6 ^e	Fr, Al, Ma, Cg		▲			▲			▲
Genève	2 ^e	Fr, Ma			▲			▲		▲
	4 ^e	Fr, Al, Ma			▲			▲		▲
	6 ^e	Fr, Al, Ma			▲		▲	▲		▲
	7 ^e	Fr, Al, Ma, La			▲			▲		▲
	8 ^e	Fr, Al, Ma, An, La, Sc			▲			▲		▲
	9 ^e	Fr, Al, Ma, An, La, Sc			▲			▲		▲
Jura	6 ^e	Fr, Al, Ma	▲	▲	▲	▲				▲
Neuchâtel	6 ^e	Fr, Al, Ma	▲		▲	▲				▲
Valais	4 ^e	Fr, Ma			▲			▲		▲
	6 ^e	Fr, Ma			▲			▲		▲
	8 ^e	Fr, Al, Ma			▲			▲		▲
	9 ^e	Fr, Al, Ma			▲			▲		▲
Vaud	2 ^e	Fr			▲					▲
	4 ^e	Fr, Ma			▲					▲
	6 ^e	Fr, Ma	▲	▲			▲			▲
	8 ^e	Fr, Ma			▲					▲

Disciplines évaluées: Français (Fr), Mathématiques (Ma), Allemand (Al), Culture générale (Cg), Anglais (An), Latin (La), Sciences (Sc)

Moment de l'évaluation: 1 = 1^{er} trimestre, 2 = 2^e trimestre, 3 = 3^e trimestre

Utilisation des résultats: 1 = Eléments comptant dans l'orientation au secondaire
 2 = Eléments indicatifs dans l'orientation au secondaire
 3 = Eléments comptant dans l'évaluation certificative
 4 = Eléments indicatifs dans l'évaluation certificative
 5 = Analyse des résultats en vue du monitoring du système scolaire

FIGURE 12 – Epreuves cantonales romandes en 2006-2007 (Ntamakiliro & Tessaro, 2010, p. 174)

Sur la base de ces figures et des commentaires faits par leurs auteurs de celles-ci, nous proposons les constats suivants :

- *Branches*. Les EC évaluent surtout des branches dites prioritaires (mathématiques, français et allemand).
- *Fréquence dans la scolarité*. Le nombre d'EC dans la scolarité varie beaucoup entre les cantons. Les cantons de Genève et Neuchâtel prévoient par exemple de nombreuses épreuves (inscrites d'ailleurs dans leur loi scolaire).
- *Fonctions et utilisation des résultats*. Toutes les épreuves ont un double rôle. Elles sont à la fois utilisées pour l'évaluation des apprentissages des élèves (rôle pédagogique), mais aussi pour le pilotage des systèmes (rôle politique). Leur rôle pédagogique peut être de deux natures, « diagnostique » visant un rôle d'information (mais pouvant entrer également dans la certification de l'élève par une note) ou « sommative » visant un rôle de bilan (pouvant entrer dans la décision de promotion ou d'orientation de l'élève). Certains cantons ont un usage plutôt sommatif des EC (GE et VD) alors que d'autres un usage plus contrasté (FR, NE et VS), voire plutôt diagnostique (BE et JU).
- *Périodes dans la scolarité*. Les EC ont lieu généralement en fin d'un cycle de deux ans (année de scolarité paire), laissant généralement les années impaires libre d'épreuves (exceptions faites de quelques épreuves diagnostiques et des épreuves de fin de scolarité secondaire). On peut constater que les EC sont plus rares au cycle 1 (1P à 2P) et leur visée est principalement diagnostique. Le cycle 2 (3P à 6P) mêle épreuves diagnostiques et sommatives. La 6P, dernière année d'école primaire, constitue une année d'épreuve sommative pour tous les cantons (sauf Berne) constituant un élément important dans le passage et l'orientation des élèves au niveau secondaire, les modalités étant différentes selon les cantons⁵⁵. Au cycle 3 (7^{ème} à 9^{ème}), les pratiques cantonales sont très variées.
- *Périodes de l'année*. Les EC peuvent être organisées en début, au milieu ou en fin d'une année scolaire (voire même en deux périodes dans certains cas). Les périodes auxquelles sont fixées les EC dépendent de leurs fonctions.

Dierendonck (2008)⁵⁶ a investigué de façon exploratoire la perception de différentes EC par les enseignants du primaire. Il arrive à des conclusions communes que nous synthétisons :

⁵⁵ Nous dirigeons le lecteur vers le descriptif de chaque procédure cantonale réalisé par Landry (2005).

⁵⁶ L'auteur a interrogé 31 enseignants du primaire de différents cantons suisses romands (Neuchâtel, Vaud et Fribourg) par *focus group* afin d'investiguer leurs perceptions de 6 épreuves externes (chaque groupe étant questionné sur une épreuve).

- *Finalités des EC.* Ces EC visent chacune des objectifs très variables. Justement, ces finalités apparaissent parfois comme floues pour les enseignants, ce qui peut entraîner une certaine méfiance quant à ces dispositifs, notamment en pensant que l'EC évalue avant tout le travail des enseignants ou de réduire leur autonomie.
- *Qualité des EC.* Le chercheur relève également des critiques envers ces dispositifs qui ne sont en fait pas des épreuves totalement externes puisque les enseignants, dans les différents cantons, sont amenés à administrer eux-mêmes les épreuves. De plus, les concepteurs et responsables manqueraient de moyens, ce qui amènerait à une moindre qualité, qualifiée d' « artisanal » par l'auteur, n'ayant pas d'encadrement scientifique.
- *Usage des EC.* Dans leur usage de ces EC, les enseignants interrogés se montrent plutôt critiques. Bien que les résultats leur permettent de situer leur classe par rapport à une norme cantonale, les enseignants déplorent le manque de feedback, utile pour leur pratique. De plus, ils expliquent les différences de résultats par des éléments dus à la composition des classes. Ils regrettent ce retour normatif et préféreraient une analyse davantage critériée (comme ce qu'on leur demande d'ailleurs de faire dans leur classe). Toutefois, selon les données de l'enquête, de telles informations ne semblent pas être la préoccupation des administrations cantonales. Les épreuves seraient surtout destinées à donner des « balises » aux enseignants pour attirer leur attention sur des contenus négligés ou pour réguler les pratiques d'enseignement et d'évaluation. Toutefois, cet effet serait faible malgré un investissement en temps et en efforts importants.
- *Perceptions des EC.* D'une part, les enseignants interrogés sont plutôt favorables à de tels dispositifs, relevant le caractère informatif de ces épreuves. Elles représentent pour eux un moyen de situer leurs pratiques par une évaluation objective et un bon complément à leur évaluation interne des élèves. D'autre part, ces enseignants relèvent les dérives de telles épreuves, craignant qu'elles prennent trop d'importance dans la carrière de l'élève ou dans l'évaluation de leurs propres pratiques. Finalement, il semble que le moment de passation de l'EC dans l'année scolaire est un élément important dans la perception des enseignants.

14. Développement de l'évaluation externe dans le canton de Fribourg

Ce point fait la description du contexte spécifique dans lequel est menée notre recherche. Il s'agit de la scolarité primaire du canton de Fribourg. Après un bref descriptif de son organisation structurelle, nous présentons les deux types d'épreuves externes des acquis des élèves qui s'inscrivent dans le cursus des élèves francophones.

14.1. Organisation de la scolarité primaire fribourgeoise

Le canton de Fribourg est composé de deux parties linguistiques (français et allemand). Toutefois, les structures principales sont identiques, régies par une même loi scolaire⁵⁷. La scolarité obligatoire du canton de Fribourg est composée de neuf années réparties en trois cycles de deux ans pour l'école primaire (1P-2P, 3P-4P et 5P-6P) et un cycle de trois ans pour l'école secondaire (7^{ème} à 9^{ème})⁵⁸. L'école enfantine (une année avant l'entrée au primaire) n'est pas obligatoire. Durant ces six années d'école primaire, l'enseignement de toutes les branches est assuré par un enseignant "généraliste" qui est en charge de sa classe durant un cycle de deux ans (ou par un duo d'enseignants travaillant à temps partiel). A la fin de l'école primaire (6P), les élèves sont orientés au secondaire I (cycle d'orientation) dans trois filières (exigences de base, générale, pré-gymnasiale) qui vont déterminer leur parcours dans l'école post-obligatoire (gymnase, maturité professionnelle, etc.).

La scolarité primaire fribourgeoise prévoit deux types d'évaluation externe des acquis des élèves : les épreuves cantonales de référence de fin de cycle et les épreuves cantonales de la procédure de pré-orientation (PPO). Il s'agit d'épreuves cantonales qui sont passées par les élèves en fin de cycle, en 2P, en 4P ou 6P⁵⁹. Elles sont créées et administrées par le Service de l'enseignement obligatoire de langue française (SEnOF), organe de la Direction de l'instruction publique, de la culture et du sport (DICS). Alors que l'épreuve cantonale de 6P a une histoire ancienne, appelée Test d'aptitudes et de connaissances (TAC) jusqu'en 2005, les épreuves de 2P et 4P sont plus récentes, ayant été graduellement mises en place dès 2005 (Ntamakiliro & Tessaro, 2010a). Nous en faisons tout d'abord

⁵⁷ Lors de la rédaction de la thèse (2014), une nouvelle loi scolaire est en cours de consultation. Nous nous référons à la loi en vigueur avant cette date.

⁵⁸ Nous nous référons aux structures et à la terminologie en vigueur au moment de la récolte de données de la thèse (2011-2012), les réformes structurelles HarmoS n'étant pas encore entrées en vigueur (nouveau plan d'études, nouvelle organisation structurelle).

⁵⁹ Notre intérêt se porte sur les épreuves externes de la scolarité primaire. Les épreuves du secondaire (7^{ème} à 9^{ème}) ne font pas l'objet de notre travail car, au moment de notre enquête, elles ne sont pas externes (tous les enseignants participent tout ou en partie à sa création).

une description sur la base d'informations fournies par le SEnOF (non publié) et les descriptions qu'en ont fait Dierendonck (2008) et Ntamakiliro et Tessaro (2010a). Ensuite, nous relevons les principales différences et nous en proposons une catégorisation selon les typologies proposées dans le cadre théorique (point 5.1. ; p.43-45).

14.2. Présentation des différentes épreuves externes du canton de Fribourg

14.2.1. Epreuves cantonales de référence de fin de cycle

TABLEAU 3 – Description des épreuves cantonales de référence de fin de cycle

Années de scolarité	Les épreuves ont lieu à la fin de chaque cycle de la scolarité primaire : 2P et 4P
Contenus évalués	Les épreuves évaluent une branche par année en alternance : mathématiques, allemand ou français. Les contenus évalués représentent une sélection de compétences inscrites dans les référentiels du cycle.
Période de l'année	Les épreuves ont lieu en fin d'année scolaire (juin).
Statut	La passation des épreuves est obligatoire ou facultative en fonction des visées spécifiques du SEnOF.
Finalités	<p>Les objectifs du SEnOF pour ce type d'épreuve sont multiples :</p> <ul style="list-style-type: none"> - Mettre à disposition des enseignants des outils d'évaluation en lien avec le plan d'études ainsi qu'avec les moyens d'enseignement. - Harmoniser progressivement les exigences de l'enseignement et de l'évaluation dans le canton. - Etablir un bilan du degré d'acquisition des connaissances et compétences développées au cours de la scolarité. - Donner aux enseignants, ainsi qu'aux élèves et à leurs parents, une référence extérieure au travail de la classe. - Tirer parti du bilan effectué en complétant les informations à transmettre au titulaire qui recevra les élèves dans les degrés d'enseignement suivants. - Permettre au SEnOF et à l'encadrement pédagogique (inspecteurs et collaborateurs pédagogiques) de récolter des informations en vue de l'analyse, de la régulation et du pilotage de l'enseignement.
Résultats	<ul style="list-style-type: none"> - Les résultats individuels sont communiqués par l'enseignant aux élèves (et év. aux parents) sous forme d'une appréciation (2P) ou d'une note (4P). Les épreuves sont intégrées dans le dossier d'évaluation des élèves. - L'analyse des résultats se fait prioritairement par l'enseignant. L'épreuve complète son évaluation interne, en lui permettant notamment de situer les résultats de ses élèves par rapport à une norme cantonale (interprétation normative). En fonction des visées spécifiques, les résultats sont pris ou non en considération dans l'établissement du bulletin scolaire (les épreuves ne servent pas à fonder une décision de promotion de l'élève à elles seules). - Les résultats globaux de l'épreuve et par classe sont analysés par le SEnOF

	(inspecteurs et responsables pédagogiques) dans un but de pilotage.
Modalités	<ul style="list-style-type: none"> - La conception des épreuves est réalisée par un groupe d'enseignants du cycle sous la responsabilité des responsables de branche (responsables pédagogiques). Les épreuves sont validées par une commission formée d'inspecteurs et de responsables pédagogiques. Les épreuves sont testées par un petit échantillon d'élèves (aucune analyse statistique n'est effectuée, si ce n'est des taux de réussite). - Les enseignants font passer les épreuves à leurs élèves dans leur classe. Le jour, la durée et les conditions de passation sont communs à toutes les classes (un livret de consignes est distribué aux enseignants). - Les épreuves sont corrigées par l'enseignant. Les critères de correction sont communs (selon un livret de corrections). - Les résultats des élèves (anonymes) sont envoyés par les enseignants au SEnOF. - Les anciennes épreuves sont archivées sur une plateforme accessible aux enseignants.

Selon l'enquête exploratoire de Dierendonck (2012)⁶⁰, les enseignants fribourgeois interrogés sont « plutôt favorables » à ces épreuves de référence. Ils relèvent leurs qualités (évaluation objective, modèle pour leurs pratiques d'évaluation) mais aussi certaines limites (passation trop tardive, difficultés à intégrer les résultats dans leurs propres observations, manque de feedback quant aux résultats).

14.2.2. Epreuves cantonales de la procédure de pré-orientation (PPO)

TABLEAU 4 – Description des épreuves cantonales PPO

Année de scolarité	Les épreuves ont lieu à la fin de la scolarité primaire : 6P
Contenus évalués	Les épreuves évaluent chaque année les branches suivantes : mathématiques, allemand, français et environnement ⁶¹ . Les contenus évalués représentent pour chaque branche une sélection de compétences inscrites dans les référentiels du cycle.
Période de l'année	Les épreuves ont lieu en mars.
Statut	La passation des épreuves est obligatoire pour toutes les classes.
Finalités	Les objectifs du SEnOF pour ce type d'épreuve sont multiples : <ul style="list-style-type: none"> - Orienter les élèves dans les filières du secondaire I⁶².

⁶⁰ L'enquête questionne un groupe d'enseignants fribourgeois (le nombre n'étant pas précisé spécifiquement) au sujet de leur perception de l'épreuve cantonale de mathématiques de 4P sous la forme d'un *focus group*.

⁶¹ L'environnement comprend les sciences naturelles, l'histoire et la géographie.

⁶² La décision d'orientation est prise sur la base d'indicateurs qui décrivent premièrement le degré d'acquisition des connaissances et des compétences de l'élève (les résultats de l'élève au premier semestre de la 6P complétés par ceux de

	<ul style="list-style-type: none"> - Mettre à disposition des enseignants des outils d'évaluation en lien avec le plan d'études ainsi qu'avec les moyens d'enseignement. - Harmoniser progressivement les exigences de l'enseignement et de l'évaluation dans le canton. - Etablir un bilan du degré d'acquisition des connaissances et compétences développées au cours de la scolarité. - Donner aux enseignants, ainsi qu'aux élèves et à leurs parents, une référence extérieure au travail de la classe. - Permettre au SEnOF et à l'encadrement pédagogique (inspecteurs et collaborateurs pédagogiques) de récolter des informations en vue de l'analyse, de la régulation et du pilotage de l'enseignement.
<p>Résultats</p>	<ul style="list-style-type: none"> - Les résultats individuels sont communiqués aux élèves et aux parents sous forme d'une note. - Les résultats sont pris en compte pour l'orientation des élèves dans les filières du secondaire I. Elles confirment ou non l'évaluation interne de l'enseignant. - Les résultats de chaque élève et de chaque classe sont collectés et analysés par les inspecteurs et les directeurs de l'école secondaire pour la décision d'orientation. Les résultats sont notamment comparés à ceux de l'évaluation interne de l'enseignant (notes du premier semestre de 6P). - L'analyse des résultats se fait prioritairement par l'enseignant. Les résultats complètent son évaluation interne. Ils lui permettent notamment de situer les résultats de ses élèves par rapport à une norme cantonale (interprétation normative). - Les résultats globaux de l'épreuve et par classe sont analysés par le SEnOF (inspecteurs et responsables pédagogiques) dans un but de pilotage.
<p>Modalités</p>	<ul style="list-style-type: none"> - La conception des épreuves est réalisée par un groupe d'enseignants du cycle sous la responsabilité des responsables de branche (responsables pédagogiques). Les épreuves sont validées par une commission formée d'inspecteurs et de responsables pédagogiques. Les épreuves sont testées par un petit échantillon d'élèves. (aucune analyse statistique n'est effectuée, si ce n'est des taux de réussite). - Les enseignants font passer les épreuves à leurs élèves dans leur classe. Les jours (deux jours), la durée et les conditions de passation sont communs à toutes les classes (un livret de consignes est distribué aux enseignants). - Les épreuves sont corrigées par d'autres enseignants du même cycle (l'enseignant corrige une petite partie pour ses propres élèves). Les critères de correction sont communs (selon un livret de corrections). - Les anciennes épreuves sont archivées sur une plateforme accessible aux enseignants.

l'épreuve cantonale) et deuxièmement son attitude face aux apprentissages scolaires (évaluée par l'enseignant, les parents et l'élève) (DICS, 2005).

Selon l'enquête exploratoire de Dierendonck (2012)⁶³, les enseignants fribourgeois interrogés sont plutôt partagés quant à leur perception de l'épreuve cantonale PPO⁶⁴. En effet, ils soulignent la qualité des épreuves et leur utilité afin de situer les résultats de leurs élèves par rapport à une norme. Toutefois, il ressort de leur témoignage que l'épreuve PPO leur fait subir une certaine pression. En effet, ces enseignants disent se sentir évalués eux-mêmes, leur évaluation interne étant confrontée aux résultats de leurs élèves à l'épreuve externe. Cette pression est exacerbée par le fait que l'épreuve ait lieu tôt dans l'année. L'épreuve PPO les contraint également à une planification très serrée des apprentissages, le rythme des élèves ne pouvant de ce fait pas être toujours respecté. Ces enseignants regrettent également que les barèmes des épreuves soient fixés selon une interprétation normative a posteriori (pour permettre une répartition des élèves dans les filières). Finalement, ils sont conscients de certaines dérives, la réussite des élèves pouvant devenir la priorité de certains.

14.3. Différences et catégorisation des épreuves cantonales fribourgeoises

Afin de comparer les deux types d'épreuves fribourgeoises, nous présentons un tableau récapitulatif qui intègre les critères de classification d'une évaluation de Rogiers (2004), que nous avons complétés. Les différences sont relevées en gras.

TABLEAU 5 – Comparaison des EC du canton de Fribourg

critères	Épreuves cantonales de référence	Épreuves cantonales PPO
Modalités		
<i>objet</i>	acquis des élèves (une branche par an)	acquis des élèves (quatre branches)
<i>type de recueil</i>	sommatif	sommatif
<i>moment</i>	fin de cycle de 2P et 4P (juin)	en cours de 6P (mars)
<i>conception</i>	externe	externe
<i>Passation et correction</i>	standardisée (parfois facultative)	Standardisée (obligatoire)
<i>correcteur</i>	interne	externe (et interne pour une petite part)
Fonctions		
<i>Pour l'élève</i>	Inform er les élèves (et leurs parents)	orienter les élèves
<i>Pour l'administration</i>	harmoniser les pratiques, piloter le système	Orienter les élèves , harmoniser les pratiques, piloter le système

⁶³ L'enquête questionne un groupe d'enseignants fribourgeois (le nombre n'étant pas précisé spécifiquement) au sujet de leur perception de l'épreuve cantonale PPO de français et de mathématiques sous la forme d'un *focus group*.

⁶⁴ Ils relèvent d'ailleurs que le dispositif d'orientation du canton n'a été plébiscité que par une courte majorité des enseignants fribourgeois lors d'un sondage.

<i>Pour l'enseignant</i>	bilan des compétences des élèves, comparaison à une norme extérieure	bilan des compétences des élèves, comparaison à une norme extérieure
Résultats		
<i>communication</i>	aux élèves, aux parents, à l'inspecteur	aux élèves, aux parents, à l'inspecteur et au directeur de l'école secondaire
<i>Usage pour l'élève</i>	prise en compte ou non dans le bulletin scolaire (par l'enseignant)	prise en compte pour l'orientation des élèves dans les filières du secondaire
<i>Usage pour l'enseignant</i>	situer les résultats de sa classe et de ses élèves par rapport à une norme cantonale	situer les résultats de sa classe et de ses élèves par rapport à une norme cantonale
<i>Usage pour l'administration</i>	pilotage du système (par l'inspecteur)	pilotage du système (par l'inspecteur), comparaison des résultats avec ceux de l'évaluation interne (par l'inspecteur et le directeur de l'école secondaire)
<i>analyse</i>	interprétation normative	interprétation normative

Les deux types d'épreuves externes inscrites dans la scolarité primaire du canton de Fribourg se différencient par certaines modalités (années scolaires, période de l'année, contenus évalués), mais c'est principalement dans leurs finalités et l'usage de leurs résultats qu'ils diffèrent. Si ces épreuves servent toutes deux au pilotage du système (harmonisation des pratiques, pilotage), l'importance de leurs résultats est différente pour les élèves et les enseignants :

- *Épreuves cantonales de référence.* Les résultats des épreuves de référence servent principalement à informer les enseignants, les élèves et leurs parents. Leurs résultats n'ont pas (ou peu) d'incidence sur le cursus de l'élève (les résultats sont anonymes), ni sur la validation des pratiques des enseignants (les résultats de l'épreuve externe ne sont pas directement comparés avec les résultats de l'évaluation interne de l'enseignant).
- *Épreuves cantonales PPO.* Les résultats des épreuves PPO jouent un rôle important pour le cursus de l'élève, déterminant en partie son orientation dans les filières du secondaire. Ils jouent aussi un rôle important pour la validation des pratiques des enseignants. Les résultats, présentés à l'inspecteur, aux directeurs d'école secondaire et aux parents, confirment ou non l'évaluation interne de l'enseignant.

Nous avons tenté de classer ces deux types d'épreuve en référence aux typologies présentées dans le cadre théorique⁶⁵.

TABLEAU 6 – Classement des EC fribourgeoises selon les typologies théoriques

sources	Épreuves cantonales de référence	Épreuves cantonales PPO
Eurydice (2009)	Sommative (mais sans conséquences sur la carrière de l'élève)	sommative
Newton (2007)	<i>sommative uses</i>	<i>summative uses</i>
Lafontaine (2012)	non-certificative	certificative
Soussi & Niedegger (2012)	formative-diagnostique	certificative
Marc & Wirthner (2012)	à visée diagnostique	à visée sommative

Selon nous, les typologies présentées dans le tableau précédent ne permettent pas de distinguer clairement les deux types d'épreuves cantonales du canton de Fribourg. L'usage des termes diagnostique, formative, sommative et certificative pour les différencier ne nous paraît pas pertinent : les épreuves cantonales fribourgeoises sont toutes deux sommatives (Allal, 1991; Bloom et al., 1981; Scallon, 1988)⁶⁶ : elles permettent après une période d'apprentissage de faire un bilan qui mesure le degré de connaissances et/ou de compétences par l'élève.

Afin de catégoriser les épreuves cantonales fribourgeoises, nous préférons nous référer à notre propre typologie des épreuves externes proposée dans le cadre théorique (p.45). Selon nos catégories, ces épreuves servent toutes deux à l'évaluation DES apprentissages (contrôle des apprentissages), ayant toutes deux une visée sommative : elles servent à faire un bilan qui mesure le degré d'acquisition des objectifs par les élèves après une période d'apprentissage. Elles évaluent les objectifs du plan d'études à la fin des cycles d'apprentissage (1-2P ; 3-4P ; 5-6P). Elles sont donc en cela semblables. Les caractéristiques qui les diffèrent sont leurs finalités, l'utilisation de leurs résultats et leurs conséquences sur le cursus des élèves et sur la validation des pratiques des enseignants. Nous proposons donc de les caractériser par leurs finalités (dont découlent leurs usages) et donc de retenir, dans ce travail, les termes d' « **épreuve de référence** » pour les épreuves cantonales de référence de

⁶⁵ Les différentes typologies de l'évaluation externe sont l'objet du point 5.3.2 (p.42-44). Le classement réalisé par Tessaro & Ntamakiliro (2010) n'a pas été pris en considération pour les raisons exposées dans ce point.

⁶⁶ Les fonctions de l'évaluation scolaires sont présentés au point 5.3.1 (p.41-42)

2P et 4P et d' « **épreuves d'orientation** » pour les épreuves cantonales PPO de 6P. Le tableau 5 synthétise ces éléments :

TABLEAU 7 – Classification et définition des EC fribourgeoises

<p>Épreuves cantonales de référence de 2P et 4P = <i>Épreuve de référence</i></p>	<p>L'épreuve de référence une épreuve externe des acquis des élèves à visée sommative (contrôle des apprentissages). Elle n'est à notre sens pas diagnostique, car elle donne des informations à la fin et non au début de la période d'apprentissage.</p> <ul style="list-style-type: none"> - elle fait le bilan des acquis des élèves à la fin d'un cycle d'apprentissage (2P ou 4P); - elle a parfois un statut facultatif ; - elle sert prioritairement à informer les élèves, leurs parents et les enseignants ; - ses résultats n'ont que peu (ou pas) d'importance pour la carrière scolaire de l'élève, ayant (ou non) un poids modéré dans la décision de sa promotion au cycle suivant. - ses résultats n'ont que peu (ou pas) d'importance pour la validation des pratiques des enseignants, n'entraînant pas de comparaison directe.
<p>Épreuves cantonales PPO de 6P = <i>Épreuve d'orientation</i></p>	<p>L'épreuve d'orientation est une épreuve externe des acquis des élèves à visée sommative (contrôle des apprentissages). Elle n'est à notre sens pas certificative, car elle ne donne pas directement accès à un diplôme).</p> <ul style="list-style-type: none"> - elle fait le bilan des acquis des élèves à la fin d'un cycle d'apprentissage; - elle a un statut obligatoire ; - elle sert prioritairement à l'orientation des élèves dans les filières du secondaire I ; - ses résultats ont une grande importance pour l'élève, constituant un indicateur important pour son orientation. - ses résultats tiennent une certaine importance pour la validation des pratiques des enseignants. Une comparaison des résultats externes et internes étant faite par l'inspecteur et le directeur de l'établissement secondaire (et év. par les parents).

CHAPITRE 3

PREMIÈRE PARTIE EMPIRIQUE

15. Objectifs de la première partie empirique	113
16. Méthodologie de la première partie empirique.....	114
16.1. Approche méthodologique	114
16.2. Echantillon.....	114
16.3. Recueil des données.....	115
16.4. Traitement et analyse des données	116
17. Présentation des résultats de la première partie empirique	119
17.1. Analyses de cas (analyse intra-site).....	119
17.1.1. Analyse de cas n°1 : Aurélie	119
17.1.2. Analyse de cas n°2 : Claude	122
17.1.3. Analyse de cas n°3 : Vincent.....	125
17.1.4. Analyse de cas n°4 : Chantal	129
17.1.5. Analyse de cas n°5 : Patrick.....	132
17.1.6. Analyse de cas n°6 : Romain.....	133
17.1.7. Analyse de cas n°7 : Julien.....	137
17.1.8. Analyse de cas n°8 : Ghislain.....	140
17.2. Analyse globale (analyse inter-sites).....	142
17.2.1. Perceptions des enseignants quant à l'évaluation externe	142
17.2.2. Harmonisation des pratiques d'évaluation sommative	142
17.2.3. Développement du professionnalisme des enseignants.....	145
17.2.4. Rétrécissement du curriculum	146
17.2.5. Teaching to the test.....	146
17.2.6. Encouragement à la collaboration	147
17.3. Synthèse : les effets identifiés et les variables qui les influencent	148

15. Objectifs de la première partie empirique

La présente phase empirique est qualifiée de “recherche exploratoire” car elle permet de « clarifier un problème » et de « déterminer le devis de recherche adéquat, avant de mener une étude de plus grande envergure » (Trudel, Simard, & Vonarx, 2007, p. 39). En effet, cette première analyse de données permet d’améliorer notre connaissance du phénomène étudié et de trouver de premières réponses aux questions de recherche de la thèse en vérifiant certaines hypothèses⁶⁷. Elle nous octroie également la possibilité de prendre du recul quant à nos propres pratiques d’enseignant et de laisser “parler le terrain”.

Le chercheur doit pouvoir mettre en question les comportements observés tels qu’ils se présentent d’emblée à ses yeux et tels qu’ils sont perçus et décrits par les acteurs eux-mêmes. Il doit pouvoir se ménager une position de recul et adopter une posture intellectuelle de distance critique (Albarello, 2007, p.18).

Les objectifs de cette première phase sont les suivants :

- Identifier les perceptions des enseignants quant à l’évaluation externe des acquis des élèves [QR1] ;
- Identifier les effets de l’évaluation externe des acquis des élèves sur les pratiques des enseignants [QR2] ;
- Identifier les variables qui influencent les effets de l’évaluation externe des acquis des élèves sur les pratiques des enseignants [QR3].

Les résultats de cette phase vont ainsi permettre de centrer la deuxième partie empirique de la thèse sur les principaux effets identifiés et sur les variables qui les influencent. Elle va également permettre la construction d’un outil de mesure qui permettra, dans un premier temps, de vérifier ces résultats à plus grande échelle (sur un échantillon plus large), puis de mesurer la variabilité et les relations entre les variables identifiées.

⁶⁷ Les questions de recherche et hypothèse sont présentées au point 10 (p.88-90) .

16. Méthodologie de la première partie empirique

16.1. Approche méthodologique

Cette phase exploratoire a été menée en suivant une méthodologie de recherche qualitative. L'approche phénoménologique a été choisie parmi les multiples possibilités de procéder car elle apparaît comme la plus adéquate pour remplir nos objectifs.

(L'approche phénoménologique) se donne pour objectif l'analyse compréhensive d'un phénomène ; elle recourt principalement à la recherche de modèles de sens via la récolte d'entretiens semi-directifs et via l'analyse de contenu (Albarello, 2011, p.19).

Pour Creswell (2007), la phénoménologie vise la description et la compréhension d'un phénomène vécu (une expérience) partagée par plusieurs individus. Cette méthode est utilisée dans différentes disciplines telles que la psychologie, la philosophie ou les Sciences de l'éducation. Ainsi, l'objectif de cette phase exploratoire n'est en aucun cas de comparer les résultats entre les différents individus ou entre différents contextes, ni d'en suggérer une généralisation. Le but est de décrire, puis d'expliquer un phénomène. Selon Bernard (1988), cité dans Miles & Huberman (2010), la description signifie « rendre compréhensibles des réalités complexes en les réduisant à leurs composantes. [...] (Expliquer c'est) rendre compréhensible des réalités complexes en montrant comment leurs composantes sont associées » (p.172). Il s'agit alors de créer un modèle ou une nouvelle théorie. Toutefois, pour Miles & Huberman (2010), la frontière entre description et explication ne peut pas être tout à fait nette, le chercheur se trouvant parfois simultanément dans les deux actions.

16.2. Echantillon

Les intentions de cette phase exploratoire étant de comprendre un phénomène, l'analyse s'est portée sur l'expérience d'enseignants d'un même contexte (scolarité primaire du canton de Fribourg) confrontés à une même épreuve externe des acquis des élèves (épreuve cantonale PPO)⁶⁸. Pour cette première phase, le choix s'est porté sur cette épreuve cantonale (EC) car, au contraire des autres épreuves, elle est implantée depuis longtemps, est obligatoire, a lieu toutes les années et évalue chaque année toutes les branches principales. Elle représente donc un objet bien défini et familier pour les

⁶⁸ Le contexte scolaire et cette épreuve externe sont décrits dans le point 14 (p.103-110) .

enseignants⁶⁹. Aussi, nous ne faisons pas de différence entre les branches évaluées, les enseignants étant des “généralistes” (en charge de l’enseignement de toutes les branches dans une même classe).

Un échantillon de quinze enseignants a été formé préalablement. Il s’agit d’enseignants du degré primaire ayant la charge d’une classe de 5P ou de 6P (ou classe à degrés multiples 5-6P) lors de l’entretien. Ils sont tous confrontés à l’épreuve cantonale PPO tous les deux ans (en 6P) ou chaque année pour les classes de 5-6P. Le profil des participants est varié (sexe, expérience, localisation géographique, types de classe) dans l’intention de recenser des expériences différentes. De cet échantillon, nous avons sélectionné un à un des enseignants et les avons invités à un entretien individuel. Suite aux huit premiers entretiens, le recueil des expériences s’est arrêté car nous avons jugé que les données présentaient une certaine saturation, des entretiens supplémentaires n’auraient vraisemblablement pas apporté davantage d’informations significatives.

TABLEAU 8 – Description de l’échantillon

n°	prénom fictif	âge	sexe	lieu	années d’enseignement	années en 5P ou 6P	degré (lors de l’entretien)
1	Aurélie	32	F	ville	11	7	5P
2	Claude	53	M	ville	33	32	6P
3	Vincent	27	M	ville	3	2	6P
4	Chantal	42	F	ville	13	13	5P
5	Patrick	28	M	campagne	2	2	5-6P
6	Romain	34	M	campagne	14	14	6P
7	Julien	52	M	ville	29	26	6P
8	Ghislain	51	M	campagne	31	8	5-6P

16.3. Recueil des données

Les enseignants sélectionnés, tous volontaires, ont participé à des entretiens semi-directifs enregistrés d’une durée d’une heure et demie environ.

Les entretiens semi-directifs sont menés sur la base d’un guide d’entretien constitué de différents thèmes-questions préalablement élaborés en fonction des hypothèses. Un guide d’entretien comprend généralement

⁶⁹ Les responsables administratifs de l’épreuve cantonale PPO (3 responsables pédagogiques de branche et le responsable de la coordination des épreuves) ont également été interrogés. Ceci a permis d’interroger la procédure, les modalités et les finalités des épreuves, sans toutefois en faire une analyse.

une douzaine de thèmes-questions qui, sauf exception à justifier, seront abordés dans un ordre à chaque fois identique (Albarello, 2007, p.69).

Les entretiens ont été menés par le chercheur à l'aide d'un canevas d'entretien (annexe A ; p.250-253). Le canevas d'entretien a été construit dans le but de répondre aux objectifs de cette phase de la recherche. Il est composé de neuf thèmes-questions, incluant des questions spécifiques de relance prévues afin de traiter le thème dans son ensemble. Les thèmes-questions abordés sont les suivants, leur ordre ayant pu être respecté dans chaque entretien :

1. Profil général de l'enseignant ;
2. Conceptions et pratiques de l'évaluation scolaire ;
3. Expériences et perceptions des épreuves externes en général et de l'épreuve cantonale PPO ;
4. Effets de l'épreuve cantonale PPO sur les pratiques de gestion du curriculum ;
5. Effets de l'épreuve cantonale PPO sur les pratiques d'enseignement-apprentissage ;
6. Pratiques de préparation à l'épreuve cantonale PPO ;
7. Effets de l'épreuve cantonale PPO sur les pratiques d'évaluation ;
8. Usage des résultats de l'épreuve cantonale PPO ;
9. Autres commentaires, questions éventuelles.

Il est important de noter que les effets de l'évaluation externe n'ont pas été nommés directement par le chercheur. Ainsi, pour ne pas influencer les répondants et tenter d'aborder, dans un large spectre, les principales pratiques de l'enseignement, les questions portant sur les effets de l'évaluation externe ont été séparées en trois axes qui correspondent à trois domaines de l'enseignement : a) la gestion du curriculum ; b) les pratiques d'enseignement-apprentissage ; c) les pratiques d'évaluation interne. D'autres aspects de la pratique enseignante auraient pu être explorée (par ex. relations avec les parents, relations avec les collègues) mais notre intérêt se porte surtout sur les pratiques (non pas sur les attitudes).

16.4. Traitement et analyse des données

Pour répondre aux questions de recherche de cette phase exploratoire, une analyse de contenu a été réalisée sur les données issues des entretiens (L'Écuyer, 1990; Miles & Huberman, 2010; Mucchielli, 2006).

Analyser le contenu (d'un document ou d'une communication), c'est [...] rechercher les informations qui s'y trouvent, dégager le sens ou les sens de ce qui est présenté, formuler et classer tout ce que « contient » ce document ou cette communication (Mucchielli, 2006, p.23).

Dans un premier temps, les entretiens enregistrés ont été transcrits fidèlement (à l'aide du logiciel *Hypertranscribe*) et rendus anonymes. Ce matériel textuel a permis une analyse de type logico-sémantique, c'est-à-dire une compréhension du sens du discours et non de sa forme (Mucchielli, 2006).

Notre méthode d'analyse de contenu suit les étapes proposées par L'Ecuyer (1990). Cet auteur se base sur différents ouvrages, notamment celui de Mucchielli (1979).

1. Les entretiens sont relus dans leur ensemble pour en faire une « analyse flottante ».
2. Le texte est découpé en unités de sens. Ce sont des « énoncés plus restreints possédant un sens complet en eux-mêmes ». Il peut s'agir de mots, de phrases ou même de paragraphes.
3. Les unités de sens sont classées à l'intérieur de catégories par analogie de sens⁷⁰. Grâce au classement des données, les catégories sont ensuite multipliées (nouvelles catégories) et/ou affinées (plusieurs niveaux sont créés dans une même catégorie). Cette étape a été facilitée par l'emploi du logiciel *Hyperresearch*. L'annexe B (p.254-257) présente l'arborescence finale des catégories utilisées.
4. Finalement, L'Ecuyer (1990) propose une analyse quantitative et qualitative du matériel, puis une interprétation des résultats. Selon la posture définie et les objectifs de la recherche, le choix a été fait pour cette phase de faire uniquement une analyse qualitative de ces données. Le matériel textuel réorganisé en catégories a pu être analysé en suivant la méthode de Miles & Huberman (2010) qui comprend deux étapes :
 - *analyse de cas (analyse de cas intra-site)*. Cette phase a pour but d'identifier les effets de l'évaluation externe pour chaque enseignant séparément. Pour chacun, une progression analytique est réalisée comme proposée par Rein & Schon (1997) cités par Miles & Huberman (2010). Il s'agit en effet tout d'abord de « relater un premier « récit » sur une situation spécifiée (que s'est-il passé, et ensuite que s'est-il passé ?) pour construire une « carte » (qui formalise le récit, qui localise les variables-clés) et pour élaborer une théorie ou modèle (sur la façon dont les variables sont reliées, sur l'influence qu'elles ont les unes sur les autres) » (p.173). Pour chaque enseignant, une *description* sous forme de texte a été faite et illustrée de verbatim. Puis, sur la base de ce texte, un « diagramme causal » accompagné d'un texte explicatif a pu être créé

⁷⁰ Des étudiants de Master en Sciences de l'éducation de l'Université de Fribourg ont collaboré aux phases de transcription et de classification des données. Leur apport a permis de confronter les visions et de donner une certaine valeur supplémentaire aux catégories et à la classification des données (validité inter-juges). Nous remercions Anne-Loise Cerantola, Véronique Johner, Corinne Huguenot, Valérie Le, Chrystel Schwab, Luisa Goldhorn, Benjamin Gasser et Katia Boschung.

pour *expliquer* le fonctionnement de chaque cas. Cette méthode de présentation des résultats a été retenue parmi d'autres car elle répond aux objectifs de la présente recherche.

Un diagramme causal est une visualisation des plus importantes variables indépendantes et dépendantes dans une étude de terrain et de leurs interrelations. L'orientation de ces relations est déterministe plutôt que purement corrélative. On part du principe que certains facteurs exercent sur d'autres une influence directionnelle : X suscite l'apparition de Y, ou fait varier Y (Miles & Huberman, 2010, p.274).

- *analyse globale (analyse inter-site)*. Le texte et les digrammes de causalité des différents cas sont synthétisés pour en faire une analyse globale, sous forme de texte puis de méta-matrice. Cette phase d'analyse globale permet de créer des premières catégories de mode de fonctionnement et d'en ressortir les variables les plus importantes et leurs relations.

L'analyste va généralement subdiviser une nouvelle fois les données (suivant d'autres critères) et regrouper celles qui vont ensemble de façon à mieux marquer les contrastes entre familles de sites sur les variables dignes d'intérêt (Miles & Huberman, 2010, p.318).

Dans notre travail, cette étape consiste donc à faire la synthèse des effets de l'évaluation externe sur les pratiques des enseignants et des variables qui les influencent.

17. Présentation des résultats de la première partie empirique

Ce point présente les résultats de l'analyse de contenu réalisée sur le discours des huit enseignants interrogés. Elle suit la méthode présentée dans le point précédent (Miles & Huberman, 2010). L'analyse se porte tout d'abord sur chaque cas (analyse de cas intra-site), puis les données sont analysées de manière globale, en synthétisant les résultats de tous les participants (analyse inter-site).

17.1. Analyses de cas (analyse intra-site)

Chaque analyse de cas comprend :

- Un « récit » qui décrit le profil de l'enseignant (sa carrière, ses conceptions de l'évaluation scolaire), ses perceptions de l'évaluation externe et les effets de l'épreuve cantonale PPO sur ses pratiques (séparées en trois axes : a) ses pratiques d'évaluation sommative⁷¹, b) ses pratiques de gestion du curriculum et c) ses pratiques d'enseignement-apprentissage).
- Une « carte » sous forme de diagramme causal qui représente le fonctionnement de chaque cas. Lors de cette étape, les effets de l'épreuve cantonale PPO et les variables qui les influencent sont nommés. Un texte commente le diagramme.

17.1.1. Analyse de cas n°1 : Aurélie⁷²

Profil

Aurélie a onze ans d'expérience professionnelle en tant qu'enseignante. Elle a eu tout d'abord la charge de classes du cycle 3P-4P durant quatre ans, puis du cycle 5P-6P durant sept ans. Elle a toujours enseigné dans la même école d'un milieu urbain.

Cette enseignante voit l'évaluation comme un moyen de situer les élèves dans leurs apprentissages. Elle constate que ses évaluations sommatives se rapprochent du modèle de l'épreuve cantonale PPO: une évaluation peu fréquente et globale.

« Je trouve que la façon de faire qu'on a maintenant, de faire quelque chose de plus long mais plus global, ça correspond beaucoup mieux à ce qu'on enseigne et ce qu'on évalue tout au long de l'année. »

⁷¹ Lors de l'entretien, les enseignants ont été interrogés sur leurs pratiques d'évaluation en général (sans spécification). La grande majorité des réponses traitent de l'évaluation sommative, très rarement de l'évaluation formative (même en relançant la discussion sur ce sujet). De ce fait, l'analyse n'identifie pas d'effets de l'EC sur leurs pratiques d'évaluation formative.

⁷² Un prénom fictif a été attribué à chaque enseignant.

Perceptions de l'évaluation externe

Durant sa carrière, Aurélie a fait passer trois fois l'épreuve cantonale PPO à ses classes. Elle est tout à fait favorable à cette épreuve. C'est pour elle un bon moyen de valider ses pratiques. Elle remet par contre en cause la procédure qui englobe cette évaluation externe, qui met beaucoup de pression sur les enseignants (orientation des élèves et moment de passation trop tôt dans l'année).

Axe A : Les effets de l'évaluation externe sur les pratiques d'évaluation sommative

Aurélie affirme ne pas changer son mode d'évaluation sommative par rapport à l'épreuve PPO. Elle pratique déjà le même « modèle ». Elle dit avoir toujours évalué les apprentissages de ses élèves de cette façon, même en 3P et en 4P dans ses premières années d'enseignement. Aurélie est persuadée que ce modèle d'évaluation est le meilleur.

Aurélie change tout de même les modalités de passation (gestion du temps) de ses propres évaluations sommatives à l'approche des épreuves cantonales, afin d'habituer ses élèves.

« La seule chose qui peut changer c'est peut-être le temps que je donne aux élèves. J'essaie de les amener de plus en plus vers un temps maximal à respecter en vue de la PPO. Par contre, je travaille en évaluation globale à partir de la cinquième, dès que je les ai. Donc ça ne change pas grand-chose. »

Axe B : Les effets de l'évaluation externe sur les pratiques de gestion du curriculum

Aurélie estime que l'épreuve PPO a surtout des effets au niveau de sa gestion du programme de 6P. La plupart du temps, elle travaille en priorité les thèmes sur lesquels seront interrogés les élèves. Elle garde les contenus soit plus difficiles, soit moins importants pour le printemps, lorsque l'épreuve est passée. L'évaluation externe n'a en revanche pas de conséquence directe sur sa planification annuelle pour la 5P.

Aurélie estime bénéficier de suffisamment de temps pour traiter les contenus imposés avec ses élèves. Selon elle, l'EC ne l'empêche pas de traiter l'entier du programme. Par contre, elle ressent tout de même une certaine pression à achever tous les contenus avant l'examen, afin que ses élèves soient prêts. Elle est donc stressée par ce délai, ce qui lui permet moins d'activités annexes durant le premier semestre. Aussi pour certaines branches, elle déplore le manque de liberté qu'ont les enseignants dans le choix des contenus, le programme pour l'examen étant imposé. Au final, bien qu'elle ressente un certain stress, il est normal pour elle de se tenir à ce programme fixé en vue de l'épreuve PPO. De plus, elle considère que les objectifs principaux sont bien ciblés.

Axe C : Les effets de l'évaluation externe sur les pratiques d'enseignement-apprentissage

Aurélie dit ne pas enseigner différemment en vue de l'épreuve PPO. Elle n'a pas de méthode particulière de préparation si ce n'est, dans l'idéal, de prendre un temps pour des révisions et pour différencier son enseignement. Elle avoue tout de même ne pas avoir souvent ce temps à disposition. Aurélie organise toutefois quelques épreuves en blanc (deux ou trois) pour rassurer ses élèves.

Pour Aurélie, l'épreuve cantonale PPO a fait évoluer ses pratiques d'enseignement dans certains cas. En environnement, elle dit avoir modifié son enseignement vers plus de travail sur des compétences (lecture de documents, repérage, observation...) que sur des connaissances. L'EC est un très bon moyen d'amélioration de ses pratiques. Elle remet volontiers en cause ses pratiques par rapport aux résultats des épreuves. Elle prend cet examen de manière positive.

« Au niveau de l'environnement surtout je fais ça, oui. Il faut plus savoir-faire que savoir exactement. Moi, je donne l'histoire depuis plusieurs années aux 7-6P et c'est vrai que je n'ai jamais insisté sur les dates exactes à apprendre par cœur, mais plutôt sur lire un texte, ce qu'on en retient ou travailler dans le livre pour une évaluation, rechercher des documents d'après des années. Ça je pense que c'est une bonne compétence donc je le fais peut-être grâce à la PPO mais je crois que c'est une bonne chose. »

« Pour moi, ça c'est utile personnellement. Mais moi le résultat de mon voisin par rapport à ses élèves je m'en fous. Il faut le prendre chacun pour soi. C'est un bon moyen de réfléchir, voir si on est dans le tir ou pas du tout. »

En revenant sur son rôle d'enseignante face à cette épreuve, Aurélie explique qu'il est important d'informer les élèves et les parents sur ce qui les attend car elle remarque que souvent il y a une incompréhension liée à l'épreuve PPO. Selon elle, il est important de relativiser cet événement pour qu'il ne prenne pas le dessus sur les émotions des élèves. Il faut faire en sorte que les élèves donnent leur maximum sans perdre leurs moyens. Son rôle est de relativiser la pression. Elle dit aussi suivre les apprentissages de ses élèves avec plus de précaution.

Pour assurer un contenu attrayant pour les élèves, Aurélie travaille beaucoup en collaboration avec ses collègues de la même école. Elle observe que cette collaboration devient davantage importante lorsque l'épreuve PPO est en ligne de mire. Elle estime qu'en général le matériel didactique officiel n'est pas suffisant pour préparer ses élèves. Elle crée donc beaucoup de matériel et essaie d'adapter de manière optimale son matériel à sa classe du moment. Toutefois, bien que ceci prenne du temps et de l'énergie, elle trouve que c'est une bonne chose.

Synthèse du cas n°1 : Aurélie

Comme exposé dans la partie méthodologique, un digramme causal synthétise le « récit » de chaque enseignant (Miles & Huberman, 2010). Cette « carte » graphique représente le fonctionnement de chaque cas. Les effets de l'évaluation externe sur leurs pratiques identifiés dans leur discours de l'enseignant sont relevés et nommés (rectangles) ainsi que les facteurs qui les influencent (ovales), les

flèches signifiant « suscite l'apparition de, ou fait varier ». Le diagramme est accompagné d'un texte explicatif.

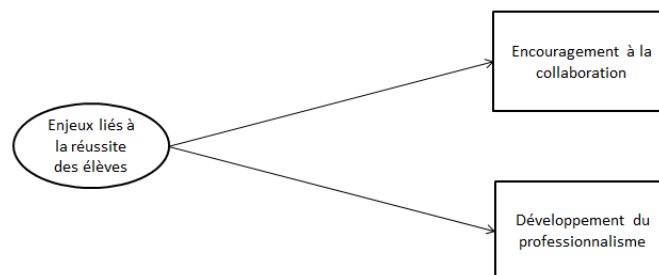


FIGURE 13 – Diagramme causal pour Aurélie

Comme l'illustre le diagramme causal, Aurélie perçoit certains effets de l'épreuve cantonale PPO sur ses pratiques. Aurélie ressent une certaine pression lors de l'année de 6P liée à la réussite de ses élèves qui, selon elle, valide ses pratiques (*enjeux liés à la réussite des élèves*). Cette pression la pousse à collaborer davantage avec ses collègues (*encouragement à la collaboration*). Elle la rend plus attentive à bien tenir son programme et à suivre davantage les apprentissages de ses élèves (*développement du professionnalisme*). Aucun autre effet n'est observé, ceci est surtout expliqué par le fait qu'elle a une grande confiance en ses pratiques, acquise notamment grâce à des expériences positives d'EC (les résultats de ses élèves aux EC concordent habituellement avec ceux de l'année). Aussi, bien qu'Aurélie soit plutôt positive face à ce dispositif d'évaluation et qu'elle le considère comme un outil de qualité, ses pratiques d'évaluation sommative ne sont que peu influencées par cette épreuve. Selon elle, ses évaluations et son enseignement ressemblent déjà beaucoup au modèle prôné par l'EC.

17.1.2. Analyse de cas n°2 : Claude

Profil

Claude est un enseignant expérimenté (33 années). Il a toujours enseigné dans le cycle 5P-6P dans la même école de milieu urbain.

Pour Claude, l'évaluation sert surtout à l'élève afin qu'il puisse situer ses apprentissages. Par contre, pour lui, elle ne sert pas ou peu à l'enseignant. Claude dit ne pas souvent remettre ses pratiques d'évaluation en question, ni les modifier. Ce sont plutôt les élèves qui doivent s'adapter.

Perceptions de l'évaluation externe

Claude a un avis négatif assez tranché sur tous les types d'évaluation externe, elles sont pour lui des moyens de contrôler les enseignants. De plus, il trouve que l'épreuve cantonale est inutile pour son travail d'enseignant. Elle le stress beaucoup. Il emploie le mot « pression » à maintes reprises. Il trouve cette pression mise sur les épaules de l'enseignant inadmissible. Il regrette également l'évolution des modèles d'évaluation et remet en cause celui de la PPO qu'il ne trouve pas bon.

« Les grosses évaluations, qui durent, qui sont longues et qui durent peut-être deux jours ou les fameuses évaluations de fin de mi-semester, je ne suis pas un fan. Parce que ça veut dire que l'on met le costume de l'évaluateur et qu'on se prépare uniquement pour cette période. »

« Je trouve que quand on étudie, il doit rester quelque chose. Aujourd'hui, c'est dommage, on n'ose plus parler de connaissances. On évalue seulement la débrouillardise, trouver une réponse à partir de documents. »

« Alors très brièvement, pour moi, c'est quelque chose, la PPO, le TAC, enfin l'examen de passage, c'est quelque chose qui me hérissé le poil depuis que j'enseigne. »

Axe A : Les effets de l'évaluation externe sur les pratiques d'évaluation sommative

Claude affirme ne pas adapter ses pratiques d'évaluation sommative par rapport aux évaluations externes. Il déclare que l'épreuve PPO ne lui donne pas de nouvelles idées. Il ne veut pas changer ses pratiques qu'il trouve bonnes. Pourtant, il pense parfois qu'il devrait le faire, voyant que les résultats de ses élèves à la l'épreuve cantonale ne concordent pas aux notes de l'année. Il déclare ne pas avoir été attentif aux évolutions des manières d'évaluer les élèves. Il remarque qu'il est un peu « en retard ». Par contre, dans la passation des épreuves, « il doit », il se sent obligé de changer les conditions de passation (surtout la durée) pour habituer ses élèves à la l'examen PPO.

« Alors, à la fin du cycle je dois, je dois (il s'exclame), les préparer à avoir des évaluations longues. Par exemple, cette année on a des évaluations qui durent environ 115 minutes de français. Alors il faut qu'ils aient joué un match de 90 minutes avant le championnat. »

D'ailleurs, les résultats de ses élèves aux EC ne sont pas très importants pour lui. Les résultats des élèves à « ses » évaluations restent l'indicateur le plus important. Parfois, constatant que ces derniers ne concordent pas avec ceux de l'EC, il pense à ajuster ses pratiques d'évaluation sommative (dans le barème et la sévérité) mais ne le fait finalement jamais, restant sceptique sur la qualité de l'EC.

On remarque une exception dans son discours pour une branche spécifique : l'environnement. Il dit que ses pratiques s'en rapprochent. Mais pour lui l'évaluation cantonale n'est pas le vecteur de ce changement de pratique, ce sont plutôt la formation continue, les échanges avec les collègues et l'évolution des moyens informatiques qui l'ont fait évoluer.

« Non, je n'adapte pas mon évaluation à celle de la PPO. Et je sais que je devrais. »

« C'est vrai que la forme de mes évaluations d'environnement se rapproche un peu de celle de la PPO. »

« Mais si je regarde mes évaluations d'il y a 20 ans. C'était surtout question-réponse. Par exemple, quelle est la formule de l'eau? H₂O. Tandis que là maintenant, avec l'évolution, ce qui est demandé, le travail avec les collègues, les cours que j'ai pris, etc., tu te rends compte que la forme évolue. »

Axe B : Les effets de l'évaluation externe sur les pratiques de gestion du curriculum

Claude organise l'année scolaire de 6P en fonction de l'épreuve PPO. La 5P ne semble que peu influencée par l'EE. C'est surtout dans l'ordre du contenu à traiter que l'évaluation externe a un effet. Claude place les contenus difficiles avant la PPO pour avantager la réussite de ses élèves. Cependant, il n'est que peu influencé par l'épreuve cantonale sur sa gestion des contenus, il dit se baser plutôt sur la progression de ses élèves. Cet enseignant prévoit tout de même une période de révisions avant les examens. Toutefois, il le fait contre son gré, il dit être obligé de faire le « PPOcompatible ». Pour lui, c'est de la « déontologie », une obligation institutionnelle. Mais ce temps peut parfois être « mangé » si les thèmes ne sont pas terminés. Claude dit avoir le temps de traiter tous les contenus de programme avant et après l'épreuve cantonale, qui d'ailleurs ne l'empêche pas de faire des activités hors programme (sorties, camps).

« Donc tu vas les mettre à ton programme. Tu ne peux pas ne pas le faire. Là tu es totalement déontologique. »

« Alors moi je fais mes programmes, je les établis en fonction de l'intérêt des élèves et après en fonction de la PPO. Donc, j'adapte plutôt en fonction du bien-être de mes élèves »

« Je suis très honnête, je fais le PPOcompatible »

« Donc il y a des choses que je place volontairement proche pour qu'ils aient tous les atouts de leur côté. »

Axe C : Les effets de l'évaluation externe sur les pratiques d'enseignement-apprentissage

Claude dit ne pas être influencé par la PPO dans sa manière d'enseigner. Il essaie en tout cas d'en faire abstraction. Cependant, on peut dénoter dans son discours qu'il prend bien garde de traiter rigoureusement les thèmes qui seront évalués dans l'épreuve PPO. Mis à part cette attention particulière, Claude dit ne pas observer de différences dans sa manière d'enseigner.

« Je voudrais dire que mon attitude personnelle et les programmes ne sont pas basés sur la PPO. Je reste totalement libre, dans ma tête et ma pratique. »

« Mais, les thèmes qui seront évalués, tu vas plus..., tu vas t'assurer qu'ils sont bien acquis. Donc tu vas donner de l'importance à ça, évidemment. »

« Je fais très peu de dictée même si je sais que chaque année paire il y a une dictée à l'examen et que moi je passe les PPO les années paires. Mais ça ne m'encourage pas à faire plus de dictées. »

A l'approche de l'évaluation cantonale, Claude fait passer une ancienne épreuve PPO à ses élèves. Son but est surtout de faire découvrir les conditions de passation à ses élèves, les sensibiliser à la durée et aux types de questions. Claude dit surtout préparer ses élèves en faisant des choses différentes, lors d'activités spéciales (camps, sorties). Il veut leur donner un état d'esprit et leur changer les idées,

surtout à l'approche de l'examen. Il compare ceci à une compétition de football : les gestes techniques sont entraînés depuis longtemps, mais ce qui compte c'est la préparation mentale.

« Sinon, je fais une PPO d'entraînement, pour qu'ils voient le questionnement et des fois je prends un bout de l'épreuve de français pour qu'ils puissent voir sur quoi on peut les interroger. Mais je ne fais rien de plus. »

« C'est lors de ces moments, camps, sorties, que je prépare la PPO. »

Synthèse du cas n°2 : Claude

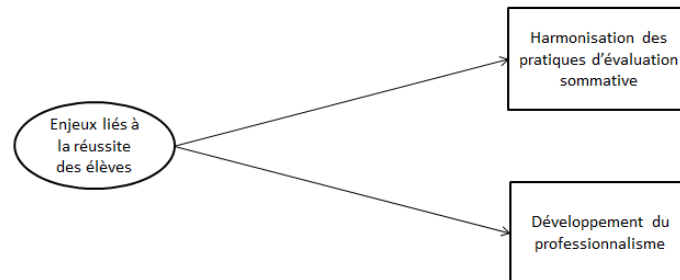


FIGURE 14 – Diagramme causal pour Claude

L'évaluation externe n'influence que peu les pratiques de Claude. C'est surtout grâce à son expérience importante et à la confiance qu'il a en ses pratiques qu'il peut rester « libre », mais aussi car il n'apprécie que peu le modèle d'évaluation de l'épreuve PPO. Toutefois, une certaine pression liée à la réussite de ses élèves l'influence (*enjeux liés à la réussite des élèves*). En effet, il doit être « déontologique » et faire le « PPOcompatible », c'est-à-dire qu'il doit faire davantage attention à ses pratiques (*développement du professionnalisme*). Cette pression le pousse aussi à modifier quelques peu ses évaluations sommatives pour que ses élèves soient « habitués » à cette forme d'évaluation (*harmonisation des pratiques d'évaluation sommative*).

17.1.3. Analyse de cas n°3 : Vincent

Profil

Vincent est un jeune enseignant. Au moment de l'entretien, il termine sa troisième année d'enseignement. Il a toujours eu la charge de classes de 5P ou de 6P dans la même école de milieu urbain.

Vincent a une vision plutôt formative de l'évaluation scolaire. Pour lui, l'évaluation sert surtout à modifier ses pratiques et à différencier son enseignement.

« C'est (l'évaluation) un moyen de fixer... de savoir où en sont les élèves, par exemple, dans un thème ce que je dois encore travailler ou non. »

« Ça (l'évaluation) me permet dans un thème de faire de la différenciation, par rapport aux besoins spécifiques des élèves, à offrir aux élèves les meilleurs moyens d'apprentissage. »

Perceptions de l'évaluation externe

Vincent a été confronté deux fois à l'épreuve cantonale PPO. Il la trouve plutôt utile pour valider et modifier ses pratiques et la définit comme un « garde-fou » dans l'orientation des élèves dans les filières du secondaire. Vincent estime que cette épreuve externe est de très bonne qualité. Il est toutefois à noter que le manque d'expérience de cet enseignant semble influencer son positionnement. L'épreuve externe lui permet de valider ses pratiques et ses prises de décision quant à l'orientation de ses élèves, mais il pense qu'au fil des années, il n'en aura plus besoin. En outre, Vincent trouve que ce dispositif engendre beaucoup de pression inutile pour les élèves et les enseignants.

« Moi je les trouve excellentes (les épreuves PPO). [...] Je trouve que, par exemple cette année, le dossier d'environnement était génial. [...] C'est quelque chose tout en couleur avec des cartes, des recherches. Et on n'évalue pas seulement les connaissances sur deux ans, mais si l'élève est capable d'aller chercher des réponses dans un dossier. »

« Je ne suis pas complètement convaincu (par l'épreuve PPO), mais je trouve qu'en ayant vécu ça cette année, ça m'a permis de me rassurer dans mon enseignement. »

« L'inconvénient c'est le stress durant l'année, ça nous met beaucoup de barrières. »

« Mais peut-être que dans dix ans, j'aurai un autre avis. Mais cette année, ça m'a permis de recadrer certains élèves, de cibler certains problèmes. Et pour eux c'était très positif. »

Axe A : les effets de l'évaluation externe sur les pratiques d'évaluation sommative

Vincent constate que ses propres évaluations sommatives ressemblent beaucoup aux épreuves PPO. Les différences ne concernent que la durée de passation et la quantité de contenus. Mais pour lui, les fondements en sont les mêmes : épreuve globale de niveau taxonomique plutôt élevé. Dans le discours de Vincent, on peut noter qu'il adapte ses évaluations sommatives selon le modèle de la PPO dans toutes les branches, surtout pour éviter toute « surprise » aux élèves au moment de la passation de l'épreuve cantonale (environnement : recherches d'informations dans des documents, observations de documents ; maths : résolution de problèmes ; français : longs textes de compréhension). Aussi, il tend vers une similitude dans la typologie des questions.

« Je me dis qu'à la PPO ils ont un dossier où ils devront chercher des informations, alors durant l'année j'essaie déjà de faire la même chose. Pas que ce soit une surprise (pour les élèves). »

« La résolution des problèmes en maths. Y a beaucoup de résolution de problèmes dans la PPO qui sont liés avec n'importe quoi, mais ils doivent résoudre, comprendre une consigne pour résoudre un problème. »

« Ils lisent souvent des grands textes, pour que lorsqu'ils reçoivent le grand texte de la PPO, ils ne soient pas surpris, et qu'ils ne tombent pas des nues. »

Vincent utilise d'anciennes épreuves cantonales PPO pour puiser des idées et des exemples pour évaluer les apprentissages de ses élèves. Il dit cependant que ce n'est pas les épreuves qui le poussent

vers ce modèle spécifique d'évaluation, mais plutôt sa formation initiale (HEP). Il prend plutôt l'épreuve cantonale comme un « outil de référence ».

« On a une valise, et il faut la remplir au fur et à mesure des expériences. Et cette année, d'avoir vu autant de PPO (anciennes épreuves), ça m'a permis de puiser dedans pour avoir des idées. »

« Donc ça ne m'a pas poussé, c'est moi qui me suis intéressé à ça. C'est juste un outil de référence. »

Les évaluations externes sont donc un outil que Vincent emploie comme modèle. Il est ouvert à celui-ci, plusieurs raisons l'expliquent. Premièrement, il a de l'estime pour ce modèle de qualité. Deuxièmement, il a besoin de ce modèle, dû à son inexpérience, pour valider ses pratiques d'évaluation sommative. Enfin, Vincent veut « conditionner » ses élèves à ce modèle PPO pour accroître leurs chances de réussite. Il déclare d'ailleurs que si les résultats de ses élèves aux épreuves cantonales n'étaient pas bons, il remettrait surtout en cause ses propres pratiques d'évaluation.

« C'est sûr à 100%. L'enseignement, on varie les formes de travail, tout ce qu'on peut, après c'est comment on évalue l'élève qui va définir ses capacités, je trouve. Si j'avais eu des énormes différences (de résultats entre ses évaluations et l'épreuve cantonale) c'est vers l'évaluation que je me serais tourné. »

Axe B : Les effets de l'évaluation externe sur les pratiques de gestion du curriculum.

Pour Vincent, l'épreuve PPO est le point charnière de l'année scolaire de 6P. Il planifie rigoureusement le programme annuel en fonction de l'épreuve. Vincent réserve également quelques semaines pour faire des révisions et des exercices de préparation spécifiques quelques semaines avant l'épreuve. Vincent s'efforce de suivre cette planification avec discipline tout au long de l'année. Il est satisfait d'avoir réussi à le faire pour ses premières expériences. Toutefois, il déclare avoir ressenti une certaine pression, craignant que tous les contenus de l'examen ne puissent être traités à temps. Sa planification très précise l'a aidé à y parvenir. Grâce à sa planification rigoureuse également, Vincent a pu traiter aussi bien les contenus évalués que ceux qui ne l'étaient pas. D'ailleurs, il ne s'est jamais empêché d'organiser des activités hors programme telles que des camps ou des sorties. Pour Vincent, cette pression retombe une fois l'épreuve passée. Cet enseignant explique que ce stress vient surtout du fait qu'il manque d'expérience, il ne connaît pas encore bien le programme de 6P.

« (En 6P) Je fais tout mon programme pour l'année, mais je sais que jusqu'à la PPO ce sera le gros morceau. [...] Il y aura beaucoup de trucs à voir. Je sais que je ne pourrai pas trop dépasser de temps, il faudra le respecter. »

« Cette année ça a monstre bien fonctionné. Je n'ai pas du tout été stressé par le temps. »

« Oui, je m'organise en fonction de ça, vu que c'est la première année que je suis en 6P. »

Axe C : Les effets de l'évaluation externe sur les pratiques d'enseignement-apprentissage

Vincent dit modifier ses pratiques d'enseignement à l'approche de l'EC. Il observe des changements entre la 5P et la 6P, mais aussi, en 6P, avant et après l'examen. Il fait, par exemple, plus de drill et est

plus « derrière » ses élèves pour s’assurer que tout est compris. Il doit être efficace dans son enseignement, il ne doit pas « perdre de temps ». De ce fait, Vincent sent que le rythme d’apprentissage est parfois trop soutenu, il ne peut pas laisser assez de temps à certains élèves. Pour cet enseignant, c’est surtout la pression du temps qui influence ces changements, mais aussi le souhait que ses élèves réussissent au mieux les épreuves.

« Il ne faut pas laisser traîner les problèmes. [...] moi je suis beaucoup plus derrière eux, à vérifier l'acquisition des connaissances. [...] J'ai vraiment pas envie qu'on passe un thème et que certains élèves aient tout oublié et que ça leur cause problème. »

Vincent observe également des changements dans son matériel didactique. Il est poussé à créer du matériel didactique pour préparer ses élèves aux épreuves. Pour ce faire, Vincent organise également des activités spécifiques (dossiers spécifiques de répétitions pour les vacances ; discussions sur comment se comporter lors de l’épreuve ; explications précises de la procédure et des conditions d’examen ; passation d’anciennes PPO à blanc en imitant les conditions de passation ; discussion des résultats et explication des difficultés rencontrées).

Enfin, Vincent dit que la PPO le pousse à collaborer davantage avec des collègues, à plus de discussions et d’échanges de matériel. Cette collaboration lui sert surtout à se rassurer, à voir s’il fait juste. Il dit d’ailleurs avoir de la chance d’avoir des collègues plus expérimentés.

Synthèse du cas n°3 : Vincent

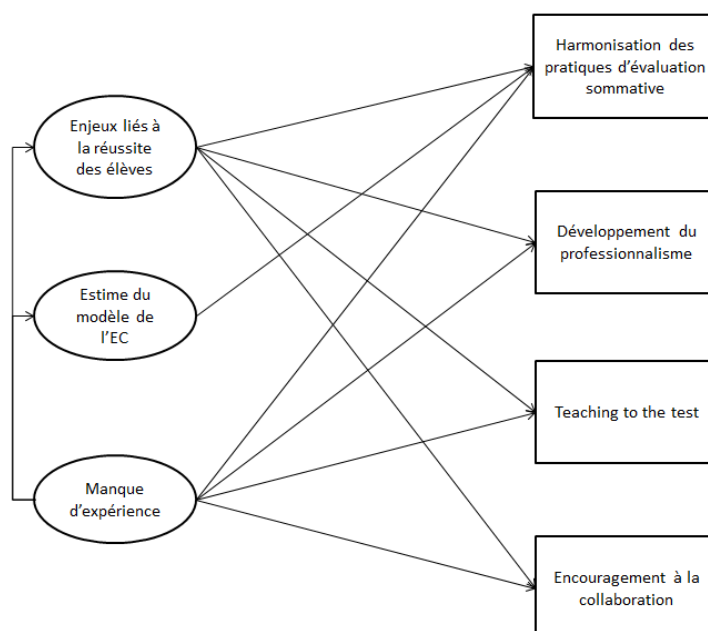


FIGURE 15 – Diagramme causal pour Vincent

Comme représenté par le diagramme causal, quatre effets de l'EE apparaissent dans le discours de Vincent. Ce jeune enseignant estime que l'EC est un outil de qualité et utile (*estime du modèle de l'EC*) qui lui sert à améliorer ses pratiques d'évaluation sommative (*harmonisation des pratiques d'évaluation sommative*). Il s'en sert surtout pour combler son manque d'expérience. Cependant, ce facteur est aussi la source d'une certaine pression, d'un stress : il veut que ses élèves réussissent et que son travail soit ainsi validé (*enjeux liés à la réussite des élèves*). Ceci le pousse à davantage d'attention sur son travail (*développement du professionnalisme*), mais aussi à des pratiques spécifiques d'enseignement (*teaching to the test*) et d'évaluation (*harmonisation des pratiques d'évaluation sommative*) pour préparer ses élèves à la réussite de l'examen et à collaborer davantage avec des collègues pour se rassurer (*encouragement à la collaboration*).

17.1.4. Analyse de cas n°4 : Chantal

Profil

Chantal est une enseignante qui a toujours enseigné dans des classes du cycle 5P et 6P, dans le même établissement scolaire de type urbain. Elle compte treize années d'expérience.

Elle voit l'évaluation comme outil nécessaire pour situer les élèves dans leurs apprentissages et pour modifier ses pratiques d'enseignement. Elle définit ses évaluations sommatives comme globales et estime qu'elles ressemblent beaucoup aux EC.

Perceptions de l'évaluation externe

Chantal est tout à fait positive envers l'épreuve PPO. Elle voudrait la maintenir mais dans une forme plus allégée. En effet, Chantal apprécie le fait que cette évaluation soit standardisée. Cette épreuve lui permet d'évaluer ses pratiques, de se « remettre à niveau », c'est d'ailleurs le point le plus apprécié par cette enseignante. Ainsi, l'EC donne une voie à suivre dans la manière de faire, mais permet aussi de tenir un certain rythme, à mener le programme de bout en bout. Cependant, Chantal regrette que ce rythme soit trop élevé. Elle se sent très stressé, elle sent une pression à devoir tout terminer à temps, pression qu'elle sent déjà lors de la 5P.

Axe A : Les effets de l'évaluation externe sur les pratiques d'évaluation sommative

Le « modèle » d'évaluation de l'EC influence les pratiques de Chantal. Cette enseignante est séduite par la forme des questions et l'approche par compétences de l'épreuve. Elle pense que l'EC montre la meilleure voie à suivre pour l'évaluation sommative des apprentissages. Les EC ont fait évoluer les

évaluations sommatives de Chantal : davantage de compréhension orale, par exemple, en français et en allemand. Elle adapte également la durée des évaluations pour entraîner les élèves.

« On cherche (dans l'EC) davantage à tester des savoir-faire plutôt que des savoirs tout court. Alors ça, c'est une bonne chose je pense. C'est ce qu'il faut faire. »

« C'est un défi pour nous car on cherche aussi à aller dans cette direction-là. »

Au fil des ans, Chantal a intégré ce modèle dans sa pratique pour trois raisons principales. Tout d'abord, elle trouve ce modèle sensé, elle le conçoit comme le meilleur modèle. Deuxièmement, elle l'intègre également pour permettre aux élèves de s'y habituer, afin d'améliorer leur réussite. Enfin, ces changements de pratiques sont également opérés sous l'influence de sa collaboration avec d'autres enseignants.

« Je sais qu'il y a un type de question qui revient à chaque fois dans toutes les PPO, je vais rendre attentif les élèves ou bien tâcher moi d'en mettre du même genre dans mes évaluations, dans mes exercices... pour qu'ils s'entraînent. »

« En français, avec les PPO je me suis rendu compte que je faisais peu de compréhension de l'oral, donc ça, ça a changé dans mon enseignement. [...] De même pour l'allemand aussi, compréhension de l'oral, on en fait un petit peu en classe, mais dans les évaluations très rarement. Donc là, y a des types de façon de faire qu'on adapte encore. »

Axe B : Les effets de l'évaluation externe sur les pratiques de gestion du curriculum

L'épreuve cantonale PPO a une grande influence sur la manière dont Chantal gère le programme de l'année. Tous les contenus peuvent être traités, mais Chantal ressent un grand stress déjà en 5P quant à la tenue du programme. Ce n'est cependant pas la réussite de ses élèves qui la met sous pression (elle est confiante quant à ses pratiques), mais bien le fait de parvenir à traiter tous les contenus à temps. Pour y parvenir, elle tient une planification précise qui doit absolument être respectée. C'est une « course à la PPO » qu'elle doit mener. L'EC détermine également l'ordre des thèmes et des contenus. D'ailleurs, cette gestion du programme rigide l'empêche de faire des activités spéciales hors programme.

« Je suis stressée au niveau du programme que je dois mener à terme mais pas l'examen en soi. »

« Pour moi, il est juste important que je sois au clair dans mon programme, que je sache où je vais pour qu'un maximum soit bouclé avant, deux semaines avant la PPO. J'aime bien me laisser le champ libre encore pour revoir ce qui doit être revu à ce moment-là. Donc c'est surtout au niveau de la planification en fait que c'est un challenge. »

« On adapte l'enseignement par rapport à ça. »

Axe C : Les effets de l'évaluation externe sur les pratiques d'enseignement-apprentissage

Pour Chantal, l'évaluation cantonale PPO est un modèle à suivre pour l'évaluation des apprentissages des élèves, mais aussi dans la manière d'enseigner, spécialement en environnement où l'on traite davantage des compétences que des connaissances. Chantal est séduite par ce modèle. Elle a donc modifié son enseignement pour qu'il y ressemble.

« Un exemple c'est en géographie, il y a très régulièrement des associations à faire entre une carte et une photo. Chose que je ne faisais pas du tout avant et que petit à petit je l'ai introduit dans mes leçons. C'est vrai que c'est intéressant »

« En français, avec les PPO je me suis rendu compte que je faisais peu de compréhension de l'oral, donc ça, ça a changé dans mon enseignement. »

Toutefois, Chantal suit ce modèle également pour habituer ses élèves à cette manière de travailler dans un but de réussite.

« Je sais qu'il y a un type de question qui revient à chaque fois dans toutes les PPO, je vais rendre attentif les élèves ou bien tâcher moi d'en mettre du même genre dans mes évaluations, dans mes exercices... pour qu'ils s'entraînent. »

De plus, Chantal observe que le stress engendré par l'abondance de contenus à traiter pour l'EC l'empêche d'enseigner de la manière dont elle voudrait (plutôt un « modèle socioconstructiviste »). Pour elle, à cause de l'EC, il est moins possible de laisser les élèves découvrir, expérimenter. La « course » décrite auparavant a pour effet qu'elle enseigne différemment. Les périodes d'apprentissage ne peuvent pas être trop longues. Elle manque parfois de temps pour un enseignement qu'elle juge adéquat. Elle se retrouve donc parfois en contradiction pédagogique.

Enfin, l'évaluation externe pousse cette enseignante à créer davantage de matériel didactique, car le matériel officiel ne suffit souvent pas. Pour ce faire, elle collabore avec d'autres enseignants de 7-6P.

« Le groupe de travail est très intéressant; en travaillant ensemble, on travaille sur des objectifs et pas sur une manière d'enseigner vu que ce sera appliqué dans 4-5 classes différentes ».

Synthèse du cas n°4 : Chantal

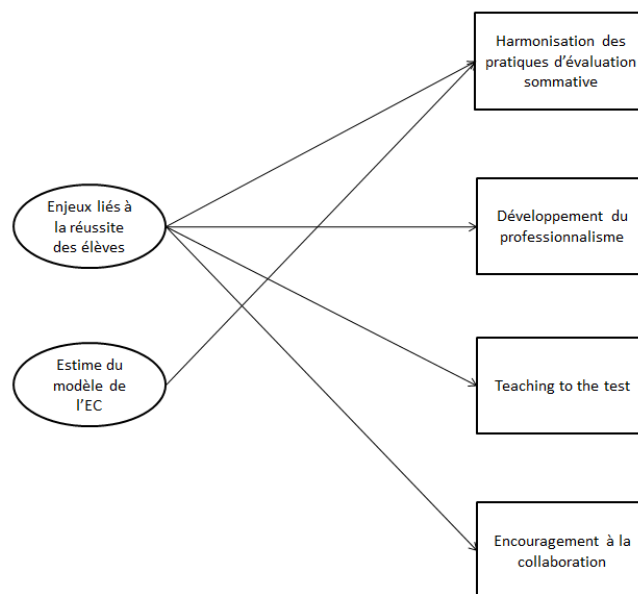


FIGURE 16 – Diagramme causal pour Chantal

L'évaluation cantonale PPO a de nombreux effets sur les pratiques de Chantal. Plusieurs raisons les expliquent. Elle estime que le modèle d'évaluation des EC est le meilleur à suivre, elle le considère comme un outil utile (*estime du modèle de l'EC*) et s'en sert pour améliorer ses pratiques d'évaluation (*harmonisation des pratiques d'évaluation sommative*). Toutefois, Chantal ressent également beaucoup de pression face à l'EC : la réussite de ses élèves la préoccupe, surtout que le temps imparti aux apprentissages avant l'examen est court (*enjeux liés à la réussite des élèves*). Ce stress l'entraîne à modifier ses pratiques d'évaluation (*harmonisation des pratiques d'évaluation sommative*) et d'enseignement (*teaching to the test*) pour préparer ses élèves à réussir l'EC. Ce facteur la rend aussi plus attentive à ses pratiques (*développement du professionnalisme*) et la pousse à collaborer davantage avec des collègues (*encouragement à la collaboration*).

17.1.5. Analyse de cas n°5 : Patrick

Profil

Patrick est un enseignant peu expérimenté qui enseigne dans une classe à double degré 5P-6P depuis deux ans, dans une école de campagne. Il est donc confronté à l'épreuve PPO chaque année pour ses élèves de 6P.

Perceptions de l'évaluation externe

Patrick est plutôt positif envers l'épreuve cantonale PPO. Il estime que c'est une bonne mesure pour mettre une pression sur les élèves et les enseignants.

« Ben moi j'ai une bonne expérience, j'aime bien parce que ça met un peu de... ça habitue un peu les élèves à avoir un peu de pression, nous aussi on a un peu de pression parce que on se dit que la classe elle doit être prête pour cet événement donc ça permet qu'on ne se repose pas sur nos lauriers. »

« Je veux dire ça a jamais tué personne cette PPO. »

Axe A : Les effets de l'évaluation externe sur les pratiques d'évaluation sommative

Patrick prend la PPO comme modèle pour ses propres évaluations sommatives. Il fait des évaluations longues, globales, voir même interdisciplinaires en environnement. Dans son discours, on peut observer qu'il a surtout adopté ce modèle dans un but de réussite de ses élèves. La PPO lui donne des idées, des exemples qu'il utilise ensuite dans ses pratiques. Il veut habituer ses élèves à un certain mode de fonctionnement.

« Oui parce que c'est obligatoire. Par exemple, pour tout ce qui est de l'environnement, il y a énormément d'analyses de schéma. Et si l'élève n'a jamais été confronté à l'analyse de schéma et qu'on lui fait l'environnement à l'ancienne où il faut tout apprendre par cœur, le jour de la PPO, il ne va pas s'en sortir. »

Axe B : Les effets de l'évaluation externe sur les pratiques de gestion du curriculum

Pour Patrick, l'épreuve PPO met une certaine pression sur sa gestion du programme, il ne doit pas « perdre » de temps. Toutefois, l'épreuve PPO ne l'empêche pas de traiter toutes les branches et d'organiser des activités hors programme (sorties, camps).

Axe C : Les effets de l'évaluation externe sur les pratiques d'enseignement

Patrick n'observe pas de changements particuliers dans son enseignement dus à l'épreuve cantonale. La seule différence se trouve dans la gestion du temps. Il sent parfois que les apprentissages ne sont pas faits de manière optimale, que la matière ne peut pas être vue en profondeur, à cause des échéances.

Patrick fait partie d'un groupe de travail qui se voit régulièrement et planifie ensemble les apprentissages de manière précise et à court terme. De ce fait, cet enseignant est rassuré. Le fait de collaborer lui permet également de varier ses évaluations. Il considère alors que ce sont déjà une sorte d'évaluation externe.

« Pour moi ça ne change rien. La PPO ça me prend deux jours c'est tout. Et le reste de l'année c'est toujours ce que je fais toujours. »

Synthèse du cas n°5 : Patrick



FIGURE 17 – Diagramme causal pour Patrick

L'épreuve cantonale PPO n'a que peu d'effets sur les pratiques de Patrick, malgré son manque d'expérience. Il dit modifier ses pratiques d'évaluation sommative (*harmonisation des pratiques d'évaluation sommative*) dans le but d'améliorer la réussite de ses élèves et ainsi de valider ses pratiques (*enjeux liés à la réussite des élèves*). Le fait de travailler en étroite collaboration avec d'autres enseignants rassure Patrick et évite d'autres effets.

17.1.6. Analyse de cas n°6 : Romain

Profil

Romain compte une quinzaine d'années d'expérience dans l'enseignement primaire. Il a toujours enseigné dans le cycle 5P-6P dans le même village de campagne.

Perceptions de l'évaluation externe

Romain est « plutôt favorable » à l'épreuve cantonale. Il affirme que l'épreuve PPO est bénéfique pour lui, elle lui donne une voie à suivre, un « garde de fou », qui lui permet de contrôler son travail et de tenir son programme. Romain pense que c'est une situation d'examen bénéfique pour l'expérience des élèves. Par contre, il déplore le trop grand stress qui touche certains élèves. Romain ressent également une certaine pression, un stress. Mais ce stress est, pour lui, dû à la densité de programme à tenir et non aux résultats de ses élèves. Finalement pour lui, ce n'est pas l'examen qui est négatif, mais plutôt ce que l'on fait des résultats : « les statistiques, les classements ».

Axe A : Les effets de l'évaluation externe sur les pratiques d'évaluation sommative

Romain constate que ses propres pratiques d'évaluation sommative ressemblent fortement au modèle d'évaluation de l'épreuve PPO. Il aimerait même que ses évaluations lui ressemblent davantage, surtout dans le fait d'évaluer plutôt des compétences que des connaissances. Il reste « insatisfait » sur ce point.

« On se rend compte qu'on aimerait évaluer des compétences, mais ce n'est pas encore ça... »

« On sent qu'on aimerait tirer vers la compétence mais que ce n'est pas toujours facile, pas toujours possible, je pense qu'on est toujours dans la connaissance mais on est souvent à la limite. »

Romain a confiance en ses pratiques d'évaluation. D'une part, il travaille en collaboration étroite avec des collègues, et d'autre part, il voit que ses pratiques et que les résultats de ses élèves correspondent bien à ceux de l'évaluation cantonale.

Romain dit que les évaluations cantonales lui donnent des idées, des exemples, mais se demande quelle est véritablement la part de responsabilité de l'EC dans cette influence. Romain voit plutôt l'évaluation comme instigatrice de questions qui viennent compléter une réflexion préexistante, Mais il ne sait pas trop bien d'où vient cet intérêt (formation, collaboration, lecture...). Finalement, Romain dit ne pas avoir adopté ce modèle pour préparer ses élèves à réussir, mais plutôt par intérêt personnel.

« Je pense qu'on se rapproche de ce questionnement-là (celui de l'épreuve cantonale). Partant du principe que ça correspond au questionnement référence. [...] Donc on va s'en inspirer, plus pour aller dans le même questionnement que pour préparer l'épreuve, pour avoir un petit peu les mêmes sensibilités, pour avoir l'impression de s'orienter vers de la compétence. »

« Mon évaluation a changé, mais maintenant quelle est la part de responsabilité de la PPO ? Je n'en sais rien. Je ne pense pas que ce soit surtout ça qui m'ait fait changer mon évaluation. Ça a sûrement influencé. [...] Elles ont pu me donner des idées. »

Axe B : Les effets de l'évaluation externe sur les pratiques de gestion du curriculum

L'évaluation cantonale a des effets sur les pratiques de gestion du curriculum de Romain. Ses pratiques ont d'ailleurs évolué avec « la répétition des épreuves ». Il se sent de moins en moins libre. Romain travaille avec des collègues du même cercle scolaire pour la planification annuelle de 6P. L'examen dicte la chronologie du programme. Les contenus de l'épreuve sont organisés de façon précise. La planification est donc rigoureuse et doit être tenue. Il « se prépare à une course », et sait qu'il ne sera certainement pas dans les temps. Il réserve justement deux semaines de réserve. Pour lui, il est tout juste possible de caser tous les contenus dans son programme, il lui faut un « chausse-pied ».

« C'est au chausse pieds et ça passe au millimètre près, on fait en sorte que ça passe. Si on avait deux semaines de plus on les utiliserait volontiers, voire trois, voire quatre. Ça va parce qu'on fait en sorte que ça convienne. »

Romain voit une différence dans sa gestion du programme avant et après l'épreuve. L'EC passée, il se sent plus libre. L'organisation de « l'après PPO » est d'ailleurs plus « aléatoire, voir anarchique ». Il se sent alors plus libre, moins stressé lors des années durant lesquelles il n'y a pas d'EC.

« On va bien organiser ce qui précède, en tâchant de mettre des semaines de réserve parce qu'on sait qu'on va être pris de cours. Et puis on organisera ensuite ce qu'il y a après l'épreuve, de manière moins... Bon il y a moins de matière, moins de notions. Je ne veux pas dire aléatoire ou anarchique mais on cherche moins de sens dans la construction qu'on donne après l'épreuve qu'avant. »

L'évaluation externe a également eu une influence sur la manière dont Romain traite la matière en classe. Auparavant, cet enseignant se permettait plus de liberté, moins de rigueur dans tous les domaines. Mais avec la PPO, il tient son programme avec rigueur, il ne se permet plus de négliger certains contenus.

« Je pense qu'au début je me détachais plus de la PPO. En y réfléchissant je pense que je, je vais employer le mot naïveté mais dans un sens pas péjoratif. Y avait des choses que je ne savais pas et je prenais plus de liberté ou j'allais plus volontiers à gauche à droite et les élèves n'étaient pas forcément perdants avec la répétition des procédures je me méfie de tout en disant ben tiens une année ça va coïncider la...et j'ai beaucoup plus de peine à laisser tomber certaines choses. Par la répétition des épreuves ça me bloque un peu dans le programme. »

Axe C : Les effets de l'évaluation externe sur les pratiques d'enseignement

Romain ressent clairement des changements dans ses pratiques d'enseignement dus à l'évaluation externe. Pour preuve, il observe une différence entre la 5P et la 6P. En 6P, il se sent plus libre.

« Je pense que ça change beaucoup sur les deux ans où j'ai les élèves. Durant la 7ème, je pense que je suis très détaché des épreuves et en 8ème au fur et à mesure qu'on s'en approche, je pense que mon enseignement est quand même fortement influencé par ce que je peux sentir de l'épreuve. »

Romain observe des modifications dans sa façon d'enseigner. En effet, Romain ressent une grande pression à devoir traiter tous les contenus dans le temps imparti. Ceci l'oblige à être très « réactif ». Il dit devoir plus individualiser, mais n'assimile pas ceci à de la différenciation. Il ne planifie pas cette

individualisation, mais agit plutôt dans l'urgence. Il ne veut pas que ses élèves « perdent du temps » dans leurs apprentissages. Il pense d'ailleurs que cette façon de faire n'est pas la meilleure pour un bon climat d'apprentissages. Il ne peut donc pas agir de la même façon que lors de la 5P, où il peut mieux organiser les apprentissages, faire de l'évaluation formative et remédier de façon différenciée. Le manque de temps est donc le facteur qui influence le plus ces changements de pratiques observés.

« A cause du temps. Il y a cette notion de temps qui... Pour moi la différenciation doit être un petit peu détachée du temps, en s'approchant, plus les semaines avancent, plus le critère temps est pris en compte. »

Romain dit ne pas préparer spécialement les élèves à l'épreuve. Il considère que sa mission est plutôt de s'assurer que tous les contenus ont été traités. De ce fait, il peut se détacher de l'importance de l'EC.

« Alors, j'essaie vraiment de ne pas les préparer à l'évaluation, mais de les préparer à des objectifs annoncés. Je pense que je ne prépare pas les évaluations. Je ne prépare pas les épreuves. Je me dis tous ces objectifs peuvent être abordés. C'est plus dans l'idée. Il faut qu'ils les aient abordés pour qu'ils ne soient pas trop déstabilisés, qu'ils évitent de se dire: ça on n'a pas vu. Donc j'ai vraiment l'impression que je prépare plutôt les objectifs que les épreuves, si on veut. »

Romain organise toutefois des épreuves en blanc de préparation. Les élèves parcourent en groupe des anciennes épreuves pour observer comment sont construits les examens. Puis, deux épreuves sont réalisées avec les modalités de passation « officielles ». Ces deux manières de procéder permettent à la classe de faire une analyse du questionnement, du type d'attentes.

Synthèse pour le cas n°6 : Romain

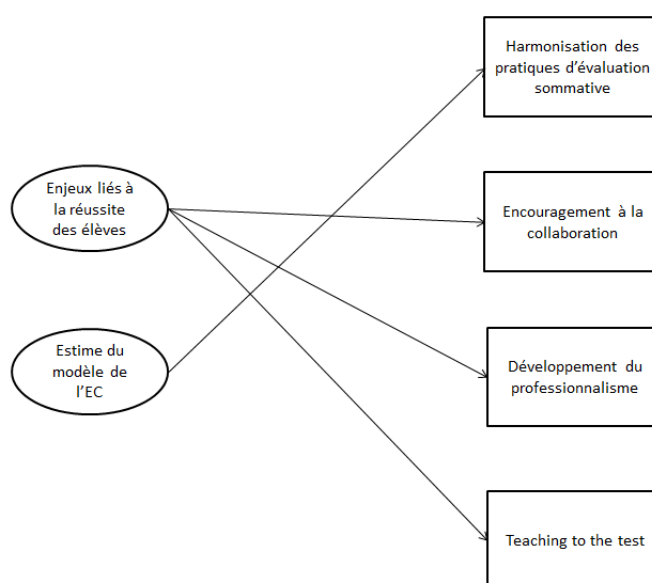


FIGURE 18 – Diagramme causal pour Romain

L'épreuve cantonale a des effets sur les pratiques de Romain. Il observe que ses pratiques sont différentes lorsque l'EC est proche. Il ressent effectivement une grosse pression à l'approche de l'EC. En effet, Romain tient à ce que ses élèves réussissent l'épreuve et que ses pratiques soient ainsi validées (*enjeux liés à la réussite des élèves*). L'EC le pousse donc à être plus attentif à ses pratiques (*développement du professionnalisme*). Cette pression, encore accrue par les délais à tenir, le pousse à enseigner de manière différente afin de préparer ses élèves à réussir l'examen (*teaching to the test*). Pour se rassurer, Romain collabore davantage avec des collègues (*encouragement à la collaboration*). Aussi, l'EC fait évoluer ses pratiques d'évaluation. En effet, il voit l'EC comme une bonne voie à suivre pour l'évaluation de ses élèves (*estime du modèle de l'EC*) et voudrait s'en rapprocher le plus possible (*harmonisation des pratiques d'évaluation sommative*).

17.1.7. Analyse de cas n°7 : Julien

Profil

Julien compte trente années d'expérience d'enseignement. Depuis le début de sa carrière, il a toujours enseigné dans des classes de 5P ou 6P dans la même école d'un milieu urbain.

Perceptions de l'évaluation externe

Julien a une longue expérience des épreuves cantonales dont il a vécu l'évolution au fil des années. Pour lui, les épreuves cantonales de 6P ont suivi une évolution positive, elles sont de moins en moins « couperets », revêtent moins d'importance pour l'orientation des élèves. Cet enseignant trouve l'EC tout à fait utile pour ses pratiques, il dit en effet qu'elles sont des outils qui servent de moyen de contrôler et de modifier ses propres pratiques. Elles lui servent aussi à maintenir une certaine pression sur les élèves tout au long de l'année.

Axe A : Les effets de l'évaluation externe sur les pratiques d'évaluation sommative

Julien prend l'épreuve cantonale PPO comme modèle pour créer ses propres évaluations sommatives. Par exemple, en environnement, il a adopté l'approche par compétence prônée par l'évaluation externe. Il ne teste plus uniquement des savoirs, mais également et surtout des savoir-faire : recherche dans des documents, observation... Aussi, en français, il cible ses évaluations sur les mêmes priorités que l'EC : expression écrite et lecture compréhension.

« Ça nous a fait beaucoup changé, ça a fait évoluer nos évaluations d'environnement. On ne teste plus uniquement des connaissances pures. On essaye toujours d'avoir des connaissances et des savoir-faire. »

« Et des fois en expression on adapte, parce que maintenant à la PPO on devrait donner de l'importance à l'expression. Avant, c'était rare qu'on testait l'expression dans une épreuve de français. »

« On les prépare à ça, à devenir des élèves observateurs, en leur disant que la réponse est sous leurs yeux mais qu'on n'a pas appris les choses, qu'on n'a pas vues concrètement. Nos évaluations font en sorte qu'on construit plus sur l'observation, sur la débrouillardise des élèves. »

Pour Julien, ses évaluations sommatives ressemblent aujourd'hui fortement au modèle des épreuves cantonales. Il désirerait même davantage de ressemblance. Le fait que Julien considère le modèle d'évaluation de l'épreuve PPO comme la voie à suivre et l'a donc encouragé à l'intégrer dans ses pratiques. Julien dit aussi agir de la sorte pour préparer ses élèves à la réussite de l'examen. Toutefois, l'influence de l'évaluation externe sur les pratiques d'évaluation sommative de Julien se fait sentir surtout sentir en 6P. En effet, Julien affirme que les évaluations qu'il construit en 5P ne ressemblent pas forcément au modèle de l'EC. Il observe donc une différence entre les deux années. On peut alors supposer que cette ressemblance est surtout due à la volonté de préparer les élèves à l'épreuve cantonale.

« Il y a de toute façon des différences, vu qu'en 5P elles ont rien à voir avec des évaluations style PPO, puisqu'en 5P on découvre, on construit la matière puis après ils sont testés sur ces nouvelles matières. »

Axe B : Les effets de l'évaluation externe sur les pratiques de gestion du curriculum

Julien planifie surtout l'année scolaire de 6P par rapport aux contenus de l'épreuve PPO. Il donne la priorité aux contenus qui y seront évalués. Cet enseignant est influencé par l'EC en 5P déjà, année où il va déjà traiter certains contenus. Julien agit de la sorte pour ne pas être pressé par le temps. De ce fait, certaines branches plus « secondaire » (musique, enbiro) vont être préteritées.

« Donc on peut déjà plus donner une orientation au travail de 5P déjà. Une orientation qui permet de souffler et d'aller tranquillement en début de 6P. Comme ça en 6P on ne doit pas régater pour être prêt au mois d'avril, pour arriver à faire tout le programme. »

Axe C : Les effets de l'évaluation externe sur les pratiques d'enseignement

Julien estime que l'épreuve cantonale PPO influence ses pratiques d'enseignement. Ces épreuves lui montrent des voies à suivre. Il modifie ses pratiques d'enseignement par rapport aux exigences de l'épreuve cantonale. Julien observe ces changements dans toutes les branches, mais surtout en environnement, où il travaille davantage sur des compétences transversales (observation, lecture, recherche) que sur des connaissances. Cependant, pour lui, ce sont plus les moyens d'enseignement qui font évoluer ses pratiques que les épreuves.

« Et je pense qu'à force de faire ces PPO, ça a une influence sur nos façons d'enseigner. »

« Je dirais que c'est plutôt les méthodes, l'introduction des nouvelles méthodes qui nous font changer nos pratiques. »

Aussi, Julien observe qu'à l'approche de l'épreuve cantonale, il est plus « derrière » ses élèves et qu'il les prépare spécifiquement à l'examen. Il organise des périodes de révisions, de drill des contenus évalués. Il dit aussi être plus vigilant sur l'atteinte des objectifs. L'épreuve PPO le force également à plus d'attention sur l'introduction de nouveaux moyens d'enseignement.

« Cette année, par exemple, c'est la première année en allemand qu'on avait la méthode "Genial". Alors on a peut-être insisté beaucoup plus sur l'allemand parce qu'on ne savait pas à quelle sauce on allait être mangé. »

Synthèse du cas n°7 : Julien

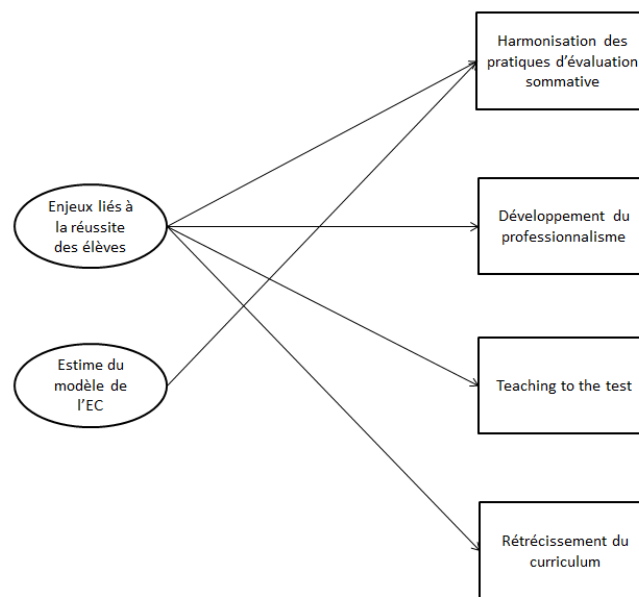


FIGURE 19 – Diagramme causal pour Julien

L'EC PPO a de nombreux effets sur les pratiques de Julien. Cet enseignant ressent beaucoup de pression face à l'épreuve cantonale, il veut que ses élèves réussissent l'épreuve et, ainsi, que ses pratiques soient validées (*enjeux liés à la réussite des élèves*). Ce facteur l'entraîne à modifier ses pratiques d'enseignement (*teaching to the test*) et d'évaluation (*harmonisation des pratiques d'évaluation sommative*) à l'approche de l'examen, à privilégier les contenus qui sont évalués (*rétrécissement du curriculum*) et à porter davantage attention à ses pratiques (*développement du professionnalisme*). Julien considère le modèle d'évaluation de l'EC comme une voie à suivre (*estime du modèle de l'EC*). Il est tout à fait positif et demandeur de tels dispositifs qui lui permettent d'améliorer ses pratiques d'évaluation sommative (*harmonisation des pratiques d'évaluation sommative*).

17.1.8. Analyse de cas n°8 : Ghislain

Profil

Ghislain est enseignant primaire depuis une trentaine d'années. Il a eu la charge de tous les degrés de la scolarité primaire, toujours dans la même école d'un milieu de campagne. Au moment de l'entretien, il enseigne depuis huit ans dans le cycle 5P-6P.

Perceptions de l'évaluation externe

Ghislain est tout à fait favorable aux EE en général. Pour lui, elles lui permettent de contrôler et de valider son travail. L'épreuve cantonale PPO lui permet de se situer, et de situer ses élèves. Elle peut également le rassurer dans ses pratiques. L'évaluation externe est, pour lui, une voie à suivre, un guide. Toutefois Ghislain, compare l'épreuve PPO à « un mal nécessaire » car elle engendre pour lui beaucoup de pression, surtout créée par les fortes attentes des parents au niveau des résultats des élèves.

« Et maintenant je suis content de ce côté-là, de pouvoir me situer. A quelque part c'est stressant, mais d'autre part ça te rassure aussi. »

« Des tests qui se font en fin de 4H, en fin de 6H, pour rendre service aux enseignants pour voir où ils doivent arriver, c'est vraiment nécessaire. »

Axe A : Les effets de l'évaluation externe sur les pratiques d'évaluation sommative

L'EC ne semble pas avoir beaucoup d'effets sur les pratiques d'évaluation sommative de Ghislain. Cet enseignant déclare clairement qu'il ne tient pas compte du « modèle PPO » pour créer ses propres évaluations. Pour lui, l'influence se fait peut-être de manière inconsciente. Il tire parfois des idées, mais plutôt sur la forme que dans le contenu ou le style de tâche à réaliser ou des exemples d'exercices. Ghislain dit que ses évaluations sommatives et les évaluations PPO ne se ressemblent pas. Il estime que les épreuves cantonales recèlent des exercices un peu « spéciaux », qu'il ne voudrait pas intégrer dans les siennes. Il observe aussi que ses propres évaluations sont davantage basées sur des savoirs et des connaissances. Cet enseignant déclare que les différentes sortes d'évaluation externe qu'il a connues l'ont surtout fait évoluer au début de sa carrière. Ghislain n'intègre donc pas le modèle d'évaluation des EC pour deux raisons : sa longue expérience et sa réticence face à ce modèle d'évaluation.

« Dans la PPO, il y a toujours un style spécial. Par exemple, cette année, l'allemand: il y avait un dessin et ils doivent produire des phrases à partir d'un dessin. C'est des exercices qu'on n'a jamais fait avant. C'est un peu spécial. »

Axe B : Les effets de l'évaluation externe sur les pratiques de gestion du curriculum

L'évaluation externe influence Ghislain dans sa gestion du programme. Il donne beaucoup d'importance au fait que les contenus soient tous vus avant l'examen. Cette influence se remarque en 5P déjà. Ghislain fait bien attention que le programme de cette année scolaire soit terminé. Il ne faut pas être en retard au début de 6P. Il planifie son programme avec une grande rigueur, et ainsi ne ressent pas trop de pression. Ghislain peut même se permettre de revoir certains contenus de 6H (environnement, allemand). Il a à cœur que tous les contenus soient vus pour préparer ses élèves à réussir l'épreuve, mais aussi à justifier son travail.

Grâce à sa gestion précise du programme, il peut traiter tous les contenus, même ceux qui ne seront pas évalués dans l'EE. Par contre, il dit ne pas pouvoir faire souvent d'activités hors programme à cause de l'épreuve cantonale PPO.

Axe C : Les effets de l'évaluation externe sur les pratiques d'enseignement

Ghislain observe clairement un changement dans ses pratiques d'enseignement à l'approche de l'épreuve. Il dit faire plus de drill, de répétitions. Il avoue aussi mettre plus de pression sur les élèves. Par contre, l'examen ne l'empêche pas d'individualiser, de différencier. En outre, il fait passer trois ou quatre anciennes épreuves pour préparer ses élèves.

« Oui, là moi je pense qu'on fait beaucoup plus de drill, beaucoup de répétitions. »

« Oui, je me rends compte parfois que je mets aussi de la pression sur les élèves. »

Synthèse pour le cas n°8 : Ghislain



FIGURE 20 – Diagramme causal pour Ghislain

L'épreuve cantonale n'a que peu d'effets sur les pratiques de Ghislain. Cet enseignant n'observe qu'un seul changement dans sa manière d'enseigner : il prépare spécifiquement ses élèves à réussir l'examen (*teaching to the test*). Cet effet est principalement dû à la pression que Ghislain ressent pour que ses élèves réussissent et ainsi que ses pratiques soient validées (*enjeux liés à la réussite des élèves*). Le fait que Ghislain ait une longue expérience et qu'il n'estime que peu le modèle d'évaluation PPO explique qu'aucun autre effet ne soit observé.

17.2. Analyse globale (analyse inter-sites)

Dans cette nouvelle phase les données des entretiens sont reprises pour en faire une analyse globale, faisant la synthèse des différents cas. Premièrement, cette analyse présente les perceptions des enseignants quant à l'EE identifiées dans le discours des enseignants. Deuxièmement, nous décrivons les cinq effets de l'évaluation externe qui ont été identifiés (harmonisation des pratiques d'évaluation sommative, développement du professionnalisme des enseignants, rétrécissement du curriculum, *teaching to the test*, encouragement à la collaboration) et les variables qui peuvent influencer ces derniers. L'analyse globale ne tient plus compte des axes de pratiques, certains effets étant transversaux. En conclusion, une méta-matrice présente la synthèse de ces résultats.

17.2.1. Perceptions des enseignants quant à l'évaluation externe

Les enseignants interrogés sont plutôt favorables à l'évaluation externe, à une exception près (Claude). Ils relèvent toutefois des inconvénients. Nous avons identifié, dans leurs perceptions des épreuves externe, deux variables :

- *Estime du modèle de l'EE.* Pour les enseignants interrogés (à une exception près), les épreuves externes sont utiles afin de valider leurs pratiques. Ce genre de dispositif leur donne des informations pour évaluer la pertinence de leurs pratiques. L'évaluation externe leur donne aussi des indications sur les contenus à traiter. Par contre, pour un seul enseignant (Claude), l'évaluation externe (en général) est totalement inutile pour son métier, ses propres pratiques lui suffisent. Enfin, tous les enseignants relèvent la bonne qualité des épreuves.
- *Enjeux liés aux résultats des élèves.* Tous les enseignants interrogés relèvent le même inconvénient qui est lié aux résultats des élèves. En effet, les enjeux liés aux résultats sont trop importants pour les élèves et leurs parents (orientation) et pour les enseignants (validation de leur pratique ; contrôle de leur travail). Les enseignants ressentent donc un certain stress. De plus, les délais très courts (lorsque l'épreuve a lieu en cours d'année scolaire et non à la fin) mettent une pression supplémentaire sur les épaules de enseignants qui doivent parvenir à traiter tous les contenus à temps.

17.2.2. Harmonisation des pratiques d'évaluation sommative

a) Définition de l'effet

L'analyse des entretiens réalisés avec les huit enseignants montre que l'épreuve cantonale PPO peut avoir des effets sur certains aspects de leurs pratiques d'évaluation sommative. Elle peut amener les enseignants à ajuster leur niveau d'exigence, notamment en confrontant les résultats de leurs propres

évaluations à ceux de l'épreuve externe. En effet, l'enseignant reçoit les résultats globaux (total des points et note) par branche de sa classe, de chacun de ses élèves et une moyenne cantonale. Toutefois, cet aspect est peu présent dans le discours des enseignants interrogés. L'effet qui est le plus observé sur leurs pratiques évaluatives concerne la conception interne de tests sommatifs (le format des tests). Les différentes épreuves externes qu'ils rencontrent au cours de leur carrière constituent un « modèle » pour évaluer les apprentissages de leurs élèves à la fin d'une séquence d'enseignement. Elles leur donnent des exemples à suivre : de nouvelles idées d'exercices ou de tâches (approche par compétences, tâches d'un niveau taxonomique plus élevé, contenus davantage globaux ou interdisciplinaires). L'EC PPO fonctionne ainsi comme un levier didactique, qui amène les pratiques des enseignants vers un modèle. Elle a donc comme effet d'harmoniser les pratiques d'évaluation sommative⁷³. Cette harmonisation est d'ailleurs une des finalités que l'administration scolaire attend de l'épreuve PPO⁷⁴. Cet effet est observé dans toutes les branches évaluées (tableau 9).

TABLEAU 9 – Leviers didactiques de l'épreuve cantonale PPO par branche

<i>Branches</i>	<i>Leviers didactiques sur les pratiques d'évaluation sommative</i>
L1 (français)	Les différents domaines du français (expression, lecture, grammaire, vocabulaire et orthographe) sont évalués de manière intégrée (ou globale) et travaillés autour du texte. L'accent est mis sur la production et la compréhension de textes.
L2 (Deutsch)	L'évaluation sommative porte davantage sur des situations de communication. Les différents domaines langagiers sont évalués (compréhension et expression orale et écrite).
Mathématiques	L'évaluation sommative porte davantage sur la résolution de problèmes.
Environnement	L'évaluation sommative est dirigée vers l'« approche par compétences » ou davantage de « savoir-faire » sont évalués par des tâches telles que l'observation, l'utilisation et la comparaison de documents.

Toutefois, l'analyse globale montre que cet effet peut prendre différentes formes, selon les motivations des enseignants :

⁷³ Il ne s'agit ici pas d'un effet de « régulation » comme définit au point 5.2.2 (p.36-37). Afin de développer leurs pratiques de conception des tests sommatifs, les enseignants n'ont ici pas besoin des résultats de l'épreuve externe (il n'y a donc pas de feedback).

⁷⁴ Les finalités de l'EC PPO sont décrites au point 14.2.2 (p.105-107)

- *Imitation du modèle.* L'enseignant ressent le besoin d' « imiter » le modèle d'évaluation de l'épreuve PPO dans le but de préparer ses élèves à sa réussite. Il utilise ces nouvelles pratiques d'évaluation sans toutefois renoncer à ses pratiques préexistantes. On remarque par exemple que certains enseignants imitent le modèle PPO uniquement les années d'épreuve cantonales ou lorsque celles-ci approchent.
- *Intégration du modèle.* L'enseignant, étant convaincu que le modèle d'évaluation de l'épreuve est le meilleur, modifie ses pratiques en y faisant référence. Ces nouvelles pratiques modifient les pratiques préexistantes.

b) *Les variables qui influencent cet effet*

Les enseignants interrogés observent des changements dans leurs pratiques d'évaluation sommative dus aux EC. Cependant, cet effet peut dépendre (dans leur forme et dans leur importance) de certains facteurs.

- *Estime du modèle d'évaluation de l'EC.* L'effet d'harmonisation des pratiques d'évaluation sommative est influencé par le regard que porte l'enseignant sur l'épreuve externe. En effet, le levier didactique est plus important quand l'enseignant trouve l'épreuve de qualité et a une certaine estime pour le modèle d'évaluation prônée par celle-ci. Ce facteur entraîne donc l'enseignant à intégrer ces pratiques, plutôt qu'à les imiter.
- *Enjeux liés aux résultats des élèves.* L'enseignant peut être amené à modifier ses pratiques d'évaluation sommative à cause des enjeux qu'il ressent face au dispositif. Ces derniers peuvent être de plusieurs natures mais sont toujours liés à la réussite des élèves : pour l'enseignant, il peut s'agir d'un souci pour le futur de ses élèves, d'une volonté de valider ses pratiques ou encore une pression liée aux délais de l'examen.
- *Expérience professionnelle de l'enseignant.* L'analyse des entretiens montre que l'expérience de l'enseignant peut avoir une influence sur cet effet. On peut supposer que les jeunes enseignants, n'ayant pas ou peu eu l'occasion de valider leurs pratiques, sont davantage demandeurs de dispositifs d'évaluation externe dans un but de formation, d'un modèle à suivre. Au contraire, certains enseignants plus expérimentés pourraient être plus réticents à vouloir changer leurs pratiques qui ont déjà fait leurs preuves.

17.2.3. Développement du professionnalisme des enseignants

a) Définition de l'effet

L'analyse des entretiens met en exergue le fait que l'EC PPO peut amener les enseignants à être davantage attentifs à leurs pratiques, l'évaluation externe entraîne donc un certain développement du professionnalisme des enseignants⁷⁵. Cet effet est observé dans plusieurs axes de pratiques. En effet, lors de l'année scolaire de l'épreuve cantonale (6P), les enseignants planifient leurs programmes avec davantage de rigueur que les autres années. Tous les enseignants interrogés le remarquent. Ils affirment faire davantage attention à la durée et à l'ordre chronologique des contenus à enseigner. Ils procèdent souvent, dans ce découpage, par ordre de difficultés. Leurs planifications sont faites de manière très précise, beaucoup plus que ce qui est fait en 5P. Elles sont même parfois travaillées en groupe. Les enseignants se sentent obligés d'être précis, les contenus imposés devant absolument être traités tous jusqu'à la date de l'examen. Ils se préparent pour une « course à la PPO », trouvant souvent que les contenus sont trop conséquents, ou que la période à disposition est trop courte. Il faudrait même un « chasse-pied » pour placer tous ces contenus « au millimètre près ». Certains enseignants observent d'ailleurs déjà cet effet sur la 5P, année durant laquelle ils sont déjà vigilants en prévision de l'épreuve. Aussi, certains enseignants disent être davantage attentifs aux apprentissages de leurs élèves plus l'épreuve approche. Ils sont d'avantage « derrière » leurs élèves, à contrôler leurs apprentissages. Il ne faut pas perdre de temps dans les apprentissages et il faut que les élèves réussissent l'examen.

b) Les variables qui influencent cet effet

Il apparaît donc dans le discours des enseignants interrogés que l'épreuve externe les amène à davantage de professionnalisme. Cependant, cet effet peut dépendre de certains facteurs.

- *Enjeux liés aux résultats des élèves.* L'épreuve externe met une certaine pression sur l'enseignant qui souhaite que ses élèves réussissent et que ses pratiques soient ainsi validées. Cette pression est exacerbée par les délais à tenir. Elle pousse l'enseignant à porter davantage d'attention et de rigueur à ses pratiques de gestion du curriculum, d'enseignement-apprentissage et d'évaluation. Pour certains enseignants, ce n'est pas une mauvaise chose, au contraire ils disent avoir besoin de cette « stimulation ». Mais d'autres sentent que cette situation n'est pas la meilleure pour créer un climat bénéfique à l'apprentissage. Certains enseignants s'étonnent et regrettent de transférer cette pression sur leurs élèves.

⁷⁵ Il est clair pour l'auteur que les enseignants font œuvre de professionnalisme même sans épreuve externe.

- *Expérience professionnelle de l'enseignant.* L'expérience professionnelle joue un rôle sur cet effet. Les jeunes enseignants, dont les pratiques ne sont pas encore été validées, peuvent se sentir davantage responsabilisés que les enseignants plus chevronnés.

17.2.4. Rétrécissement du curriculum

a) Définition de l'effet

Malgré la pression due à l'épreuve externe énoncée auparavant, l'entier des contenus du programme de l'année peut être traité par la plupart des enseignants interrogés lors de l'année de l'épreuve PPO. C'est surtout la rigueur de la planification de l'année scolaire qui permet de ne pas créer de rétrécissement du curriculum. Au même titre, ils estiment ne pas devoir s'empêcher d'organiser des activités hors programme (sorties, camps, etc.). Toutefois, il peut arriver que l'enseignant, par nécessité de temps, doive prêter quelque peu certaines branches qui ne seront pas évaluées (par exemple : musique, sport) au bénéfice des branches prioritaires. Il s'agit alors d'un effet de rétrécissement du curriculum.

b) Les variables qui influencent cet effet

Deux facteurs semblent influencer l'effet de rétrécissement du curriculum provoqué par l'épreuve externe :

- *Enjeux liés aux résultats des élèves.* La pression ressentie par les enseignants liée au temps imparti et à la réussite de leurs élèves peut expliquer que l'enseignant délaisse les contenus qui ne seront pas évalués.
- *Expérience professionnelle de l'enseignant.* Le manque d'expérience professionnelle peut amener l'enseignant à des pratiques de rétrécissement du curriculum. Les enjeux (pression du temps, réussite des élèves, validation des pratiques) sont certainement plus forts pour les jeunes enseignants.

17.2.5. Teaching to the test

a) Définition de l'effet

Certains enseignants interrogés estiment que l'épreuve externe entraîne des changements dans leurs pratiques d'enseignement. Cet effet est pourtant plutôt ponctuel, surtout présent lors des semaines précédant l'épreuve. Certains enseignants observent que plus l'évaluation externe approche, plus ils vont travailler spécifiquement pour préparer les élèves à réussir l'examen. Il s'agit d'un effet de

teaching to the test. Ils vont mettre en œuvre des méthodes d'enseignement telles que le drill, la révision, la répétition. Aussi, ils organisent des périodes spécifiques de révisions et créent du matériel didactique dans le but de préparer les élèves. Tous les enseignants interrogés utilisent dans ce même but, à des degrés différents, d'anciennes épreuves cantonales. En outre, ces pratiques ne conviennent souvent que peu aux enseignants, qui préféreraient en utiliser d'autres (découverte, expérimentation, débats...). Certains enseignants se retrouvent alors en contradiction avec leur modèle d'enseignement idéal ou habituel.

b) Les variables qui influencent cet effet

Deux facteurs semblent influencer cet effet :

- *Enjeux liés aux résultats des élèves.* L'enseignant peut être amené à préparer spécifiquement ses élèves à l'épreuve externe, car il est stressé par la réussite de leurs élèves. C'est aussi le manque de temps qui explique le recours à des méthodes didactiques moins coûteuses en temps.
- *Expérience professionnelle de l'enseignant.* Le manque d'expérience professionnelle peut amener l'enseignant à des pratiques de *teaching to the test*. Les enjeux (pression du temps, réussite des élèves, validation des pratiques) sont certainement plus forts pour les jeunes enseignants.

17.2.6. Encouragement à la collaboration

a) Définition de l'effet

Les enseignants interrogés collaborent pour la plupart déjà avec leurs collègues. Cependant, l'épreuve externe peut les encourager à collaborer davantage. La différence est peut-être qu'une collaboration différente s'installe. Il s'agit avant tout de se rassurer, de vérifier si leurs pratiques sont valides.

b) Les variables qui influencent cet effet

- *Enjeux liés à la réussite des élèves.* L'enseignant peut être amené à collaborer davantage avec ses collègues car il est stressé par la réussite de leurs élèves.
- *Expérience professionnelle de l'enseignant.* Le manque d'expérience professionnelle peut amener l'enseignant à collaborer davantage. Il veut s'assurer que ses pratiques peu éprouvées sont valides.

17.3. Synthèse : les effets identifiés et les variables qui les influencent

La méta-matrice (tableau 10) fait la synthèse des différents effets identifiés dans le discours des huit enseignants interrogés et des variables qui les influencent. Les entretiens menés avec les enseignants primaires ont permis d'identifier et de définir les principaux effets de l'EE sur leurs pratiques dans le contexte de la recherche (cette liste n'est probablement pas exhaustive). Ces effets sont déjà identifiés dans la littérature présentée dans le cadre théorique (p.53-72). Ils touchent les pratiques des enseignants tout au long du processus d'enseignement (planification, enseignement-apprentissage, évaluation sommative) mais aussi leurs attitudes face à leurs pratiques (professionnalisme, collaboration). Les entretiens ont permis d'identifier les variables qui influencent les effets cités. Ces facteurs sont liés à l'expérience professionnelle des enseignants mais surtout à leurs perceptions de l'EE. Toutefois, la matrice démontre que les enjeux liés aux résultats des élèves prennent une place très (trop) importante dans l'explication de ces effets. Ils cachent peut-être ou s'expliquent par d'autres variables importantes liés à l'enseignant (croyances des enseignants, profil personnel et professionnel) ou au type d'épreuve. Il est alors nécessaire dans la deuxième partie empirique d'investiguer plusieurs types d'épreuve, dont les enjeux liés aux résultats sont moins forts.

TABLEAU 10 – Méta-matrice des effets de l'EC PPO et des facteurs qui les influencent

Effets de l'EE identifiés	Définition synthétique de l'effet	Variables qui influencent les effets
<i>Harmonisation des pratiques d'évaluation sommative</i>	L'épreuve externe amène l'enseignant à modifier le format de ses évaluations sommatives vers le modèle de l'EE.	<ul style="list-style-type: none"> • Enjeux liés aux résultats des élèves • Estime du modèle de l'EC • Expérience professionnelle de l'enseignant
<i>Développement du professionnalisme des enseignants</i>	L'épreuve externe amène l'enseignant à faire davantage preuve d'attention et de réflexion envers ses pratiques.	<ul style="list-style-type: none"> • Enjeux liés aux résultats des élèves • Expérience professionnelle de l'enseignant
<i>Rétrécissement du curriculum</i>	L'épreuve externe amène l'enseignant à donner la priorité aux contenus qui y seront évalués.	<ul style="list-style-type: none"> • Enjeux liés aux résultats des élèves • Expérience professionnelle de l'enseignant
<i>Teaching to the test</i>	L'évaluation externe des acquis des élèves amène les enseignants à adopter des pratiques d'entraînement intensif.	<ul style="list-style-type: none"> • Enjeux liés aux résultats des élèves • Expérience professionnelle de l'enseignant
<i>Encouragement à la collaboration</i>	L'épreuve externe amène l'enseignant à collaborer davantage avec ses collègues.	<ul style="list-style-type: none"> • Enjeux liés aux résultats des élèves • Expérience professionnelle de l'enseignant

CHAPITRE 4

SECONDE PARTIE EMPIRIQUE

18. Objectifs de la seconde partie empirique.....	153
19. Méthodologie de la seconde partie empirique.....	154
19.1. Echantillon.....	154
19.2. Recueil des données.....	154
19.3. Présentation du questionnaire.....	155
19.4. Validation du questionnaire.....	156
19.5. Traitement des données.....	173
20. Présentation des résultats de la seconde partie empirique	174
20.1. Effets de l'évaluation externe.....	174
20.2. Profil personnel et professionnel des enseignants.....	178
20.3. Perceptions de l'évaluation externe.....	181
20.4. Croyances des enseignants à propos de l'enseignement.....	185
20.5. Modélisation des variables qui influencent les perceptions de l'EE.....	193
20.6. Modélisation des variables qui influencent les effets de l'EE sur les pratiques des enseignants	197

18. Objectifs de la seconde partie empirique

La première partie empirique (chapitre 3) a permis, dans le contexte de la recherche, d'identifier cinq effets de l'EE sur les pratiques des enseignants (harmonisation des pratiques d'évaluation sommative, *teaching to the test*, rétrécissement du curriculum enseigné, développement du professionnalisme, encouragement à la collaboration) et les variables qui influencent ces effets (enjeux liés aux résultats de l'EE, utilité de l'EE, expérience professionnelle). La seconde partie, en utilisant une méthodologie quantitative, permet de vérifier et d'élargir ces premiers résultats en interrogeant un échantillon plus large. Les deux approches se complètent donc pour répondre aux mêmes questions de recherche (*mixed method research*)⁷⁶.

De plus, la première partie exploratoire de la thèse investiguait les effets d'une seule épreuve externe, l'épreuve cantonale PPO. Pourtant, il a été démontré dans l'enquête exploratoire que, par la finalité de l'épreuve (orientation des élèves), les enjeux ressentis ont une influence (trop) importante sur tous ces effets. Pour contrôler ce facteur, il est alors important, dans le même contexte scolaire, d'investiguer en parallèle un autre type d'épreuve externe qui n'entraîne pas ou peu de conséquences pour l'élève et l'enseignant. Cette partie empirique investigate deux types d'épreuves externes inscrites dans la scolarité primaire du canton de Fribourg⁷⁷:

- *EE de référence* : les épreuves cantonales de référence du cycle 1-2P et 3-4P ;
- *EE d'orientation* : l'épreuve cantonale PPO du cycle 5-6P.

Les objectifs de la deuxième phase empirique de la thèse sont les suivants :

- Créer et valider un outil de mesure quantitatif (questionnaire) sur la base des éléments théoriques (chapitre 1) et empiriques (chapitre 3) ;
- Vérifier et généraliser les résultats de la première phase exploratoire, grâce à un échantillon plus large et à une méthodologie quantitative ;
- Mesurer et comparer les effets et les perceptions d'épreuves externes de différents types (de référence et d'orientation) ;
- Analyser la variance des variables (effets et facteurs d'influence) et leurs relations.
- Modéliser ces relations, afin de comprendre par quels processus se créent ces effets.

⁷⁶ Les questions de recherches et les hypothèses sont présentées au point 10 (p.88-90).

⁷⁷ Une description du contexte scolaire et de ces épreuves est faite au point 14 (p.103-107). Comme pour la première partie empirique, nous ne faisons pas de différence entre les branches évaluées, les enseignants étant des "généralistes" (en charge de l'enseignement de toutes les branches dans une même classe). Toutefois, nos données ne permettent pas de vérifier cette assumption.

19.Méthodologie de la seconde partie empirique

19.1. Echantillon

Pour atteindre nos objectifs, un premier échantillon de 500 enseignants francophones du degré primaire du canton de Fribourg a été composé avec l'autorisation du SEnOF, qui a fixé ce contingent maximal. Pour atteindre ce nombre, la constitution de l'échantillon s'est faite par la sélection de 32 établissements scolaires. Ceux-ci ont été sélectionnés sur la base des critères suivants : a) répartition équilibrée des enseignants dans les trois cycles d'enseignement (1P-2P, 3P-4P, 5P-6P) ; b) répartition équilibrée des établissements scolaires dans les arrondissements du canton de Fribourg ; c) répartition équilibrée des établissements entre zones urbaines et rurales ; d) répartition équilibrée des établissements selon la taille des effectifs d'enseignants (petits à grands établissements).

Les questionnaires de 30 écoles sur 32 ont été retournés (94%). 305 des 500 enseignants sélectionnés ont répondu, ce qui donne un taux de retour très satisfaisant de 61%. L'échantillon donne une certaine représentativité de la population des enseignants fribourgeois du degré primaire en ce qui concerne le genre (76% de femmes et 24% d'hommes). De cet échantillon, se dégagent deux groupes d'enseignants quant aux types d'épreuve externe de référence (61%) ou d'orientation (39%) (figure 21). Le groupe EE de référence est formé des enseignants faisant référence à l'EC du cycle 1P-2P (32% de l'échantillon) et à l'EC du cycle en 3P-4P (29%).

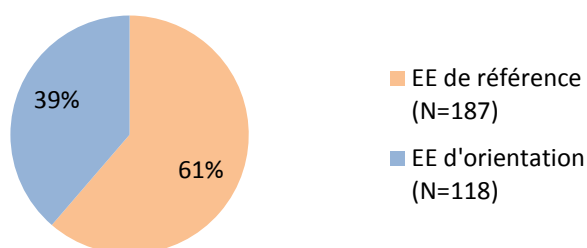


FIGURE 21 – Distribution de l'échantillon selon le type d'EE

19.2. Recueil des données

Les questionnaires ont été envoyés sous format papier à chacun des 32 cercles scolaires sélectionnés. Chaque enveloppe contenait une lettre explicative (annexe C ; p.258) et le nombre de questionnaires correspondant au nombre d'enseignants du cercle (et deux questionnaires de réserve). L'envoi a été adressé au Responsable d'Etablissement (RE) de chaque cercle scolaire. Il leur était demandé de distribuer les questionnaires aux enseignants, puis après la passation, de les récolter pour les retourner à l'aide d'une enveloppe affranchie. Afin de faciliter leur tâche, il était proposé aux RE de distribuer et de récolter les questionnaires lors d'une des réunions d'établissement qui réunissent tous les

enseignants. Il est à noter que le questionnaire n'a pas été proposé aux enseignants d'école enfantine et d'activités créatrices car ils ne sont concernés par aucune épreuve externe.

La passation des questionnaires s'est faite durant les deux premières semaines du mois de février 2012. Cette période de l'année a été choisie car elle est ni trop éloignée, ni trop proche des périodes d'évaluation externe (fin mars pour l'épreuve de 6P et juin pour les épreuves de 2P et 4P). De plus, il s'agit d'une période moins chargée pour les enseignants (les bulletins scolaires venant d'être rendus fin janvier pour le premier semestre) et qui ne comporte pas de vacances scolaires. Deux semaines ont donc été laissées aux enseignants pour remplir le questionnaire. La possibilité d'y répondre lors d'une réunion d'établissement leur a été laissée.

19.3. Présentation du questionnaire

Le questionnaire (annexe D ; p.259-265) utilisé pour mener à bien la recherche est composé de six parties comprenant chacune différentes dimensions (tableau 11). Ces variables ont été sélectionnées sur la base des résultats de la phase exploratoire (chapitre 3). Chacune d'elle est présentée par la suite.

TABLEAU 11 – Parties et dimensions du questionnaire

Parties	Dimensions
1. Perceptions de l'évaluation externe	Utilité de l'EE (UTI)
	Enjeux ressentis face à l'EE (ENJ)
	Qualité de l'EE (QUA)
	Concordance des pratiques d'évaluation interne et externe (CON)
2. Effets de l'évaluation externe sur les pratiques	Harmonisation des pratiques d'évaluation sommative (HAR)
	Teaching to the test (TTT)
	Rétrécissement du curriculum enseigné (RET)
	Développement du professionnalisme (PRO)
3. Conceptions de l'apprentissage	Encouragement à la collaboration (COL)
	Répétition (REP)
	Déséquilibre (DES)
	Apprentissage implicite (APP)
	Eradication des connaissances antérieures (ERA)
	Intégration des connaissances antérieures (INT)

	Maîtrise des savoirs (MSA)
4. Auto-efficacité personnelle	Gestion de la classe (GES) Evaluation des apprentissages (EVA)
5. Conceptions de l'évaluation scolaire	Evaluation formative (EVALFORM) Evaluation normative (EVALNORM)
6. Informations individuelles et professionnelles	Genre, âge, taux d'activité, degré d'enseignement, expérience professionnelle, type d'établissement, type de classe, forme de collaboration, formation.

Mise à part la partie 6 (informations personnelles et professionnelles), toutes les parties de ce questionnaire sont constituées d'échelles de mesure de type « Likert », les répondants devant donner leur degré d'accord par rapport à des affirmations sur une échelle allant de 1 (« pas du tout d'accord ») à 6 (« tout à fait d'accord »). Les parties 1, 2 et 6 ont été créées spécifiquement pour cette recherche alors que les parties 3, 4 et 5 (différentes croyances des enseignants) sont tirées de la littérature et ont été adaptées (Issaeva & Crahay, 2010b).

19.4. Validation du questionnaire

Avant de pouvoir analyser les données du questionnaire, il paraît indispensable de s'assurer que les outils utilisés soient pertinents et fiables. Plusieurs tests ont été menés pour ce faire. Ceux-ci ont pour objectif de confirmer si les items et les dimensions auxquelles ils se réfèrent peuvent être employés pour les analyses ultérieures. Suite à ces différents tests, certains items et/ou dimensions ont été modifiés ou supprimés pour une meilleure qualité psychométrique de l'instrument⁷⁸. Les tests effectués sont les suivants :

- *Validité des items créés.* La validité des items des échelles créées (partie 1 et 2) a été testée avant de les inclure dans le questionnaire. Six étudiants de Master (étant enseignants du degré primaire à temps partiel, à une exception)⁷⁹ ont participé à une activité de classement des items dans les dimensions selon la méthode proposée par Crocker & Algina (2008). Le but de cette procédure est de vérifier la pertinence et l'appartenance de chaque item aux dimensions prévues initialement. Elle a permis de modifier la formulation de certains items qui ne répondaient pas aux critères de validité.

⁷⁸ Toutes les modifications sont mentionnées dans ce point.

⁷⁹ Nous remercions Anne-Loise Cerantola, Corinne Huguenot, Chrystel Schwab, Benjamin Gasser, Caroline Lauper et Katia Boschung.

- *Pré-test du questionnaire.* L'entier du questionnaire a été pré-testé. Sa passation auprès de quinze enseignants du degré primaire (cinq enseignants de chacun des trois cycles de la scolarité primaire) a permis de vérifier la bonne compréhension des consignes et des items. Cette activité a été menée avec l'aide du même groupe d'étudiants. Ceux-ci ont pu assister à la passation individuelle du questionnaire et prendre note des remarques et interrogations des participants du pré-test au sujet des consignes et des items. Ce pré-test a permis de préciser quelques termes dans les items et d'améliorer les consignes, notamment en indiquant de ne pas prêter attention à la ressemblance de certains items.
- *Distribution des scores aux items.* Une analyse des données pour chaque item a été effectuée. Il s'agit des indices d'asymétrie (*Skewness*) et d'aplatissement (*Kurtosis*)⁸⁰ qui ont comme finalité de s'assurer que la distribution de chaque item et de chaque dimension soit proche de la loi normale et permet d'exclure l'éventualité de cas extrêmes (Carricano, Bertrandias, & Poujol, 2010). Pour cette recherche, ces analyses descriptives ont permis de conserver, dans un premier temps, tous les items du questionnaire et tous les répondants.
- *Construction des dimensions.* Chaque dimension est formée, *a priori*, de plusieurs items (3 à 5). Différentes analyses ont permis de vérifier la validité du construit.
 - a) Analyse factorielle exploratoire. Pour chaque partie du questionnaire, une analyse factorielle exploratoire a été réalisée pour vérifier la pertinence de la structure *a priori*. La méthode d'extraction choisie est celle de l'analyse en composantes principales (ACP), rotation Varimax. Toutes les conditions pour ce test étaient remplies pour chacune des échelles : taille de l'échantillon, type de variables et nombre d'items (Carricano et al., 2010). Aussi, pour chaque analyse, il a été vérifié que les corrélations entre items ne soient pas nulles (test de Bartlett significatif) et qu'elles soient de bonne qualité ($KMO > .60$). Après rotation, nous avons retenu les facteurs sur la base du critère de Kaiser (valeurs propres supérieures à 1) et nous avons identifié les dimensions. Pour être acceptable, l'indice de saturation doit être supérieur à $.60$ ⁸¹. Ces constructions d'items forment alors les dimensions qui sont analysées pour répondre aux questions de recherche.
 - b) Analyse de consistance interne. Suite aux analyses factorielles exploratoires, un test de consistance interne, l'alpha de Cronbach, est effectué pour vérifier l'homogénéité

⁸⁰ Le coefficient de symétrie doit être inférieur à 1 et celui d'aplatissement doit être inférieur à 1,5.

⁸¹ Les échelles négatives ayant été inversées, les indices de saturation sont positifs.

des items de chacune des dimensions. On accepte généralement un seuil minimal proche de .60 (Laveault, 2012b).

- c) Analyse factorielle confirmatoire. Pour chaque partie, une analyse factorielle confirmatoire a été menée pour valider la pertinence des structures construites par les premiers tests. Ce test permet de confirmer si le modèle est en adéquation avec les données recueillies par le biais du questionnaire. Les indices les plus fréquemment utilisés dans la littérature ont été retenus avec leur seuil usuel (Hu & Bentler, 1999): NNFI (Non-Normed Fit Index ; $>.95$), CFI (Comparative Fit Index ; $>.95$), χ^2/dl (< 3) et le RMSEA (Root Mean Square Error of Approximation ; $<.05$).

19.4.1. Partie 1 - Perceptions de l'évaluation externe

Dans le questionnaire, les participants répondent en faisant référence à l'une des trois épreuves externes de la scolarité fribourgeoise⁸². En introduction, les enseignants indiquent parmi les trois propositions l'épreuve externe à laquelle ils vont faire référence pour répondre aux items. Il leur est demandé de choisir l'épreuve cantonale à laquelle ils sont (ou ont été) confrontés le plus souvent dans leur carrière.

Cette partie du questionnaire a pour objectif de mesurer les perceptions des enseignants quant au dispositif d'évaluation externe sélectionné. Les perceptions mesurées ne sont pas exhaustives, elles sont celles qui ont été identifiées par la recherche exploratoire comme pouvant avoir une influence sur les effets de l'EE (voir tableau 10, p.149). La dimension « estime du modèle de l'EE » a été divisée en deux dimensions « utilité perçue de l'EE » et « qualité de l'EE ». La dimension « concordance des pratiques d'évaluation » a également été ajoutée. Nous présentons premièrement les dimensions prévues *a priori* dans notre questionnaire, puis les dimensions validées par les analyses et utilisées par la suite dans l'analyse des données.

a) Dimensions du questionnaire (a priori)

A priori (avant les analyses de validation), la partie 1 était composée de 12 items investiguant 4 dimensions (tableau 12).

⁸² Les épreuves cantonales de référence du cycle 1-2P ; les épreuves cantonales de référence du cycle 3-4P ; les épreuves cantonales PPO du cycle 5-6P

TABLEAU 12 – Items et dimensions de la partie 1 (a priori)

<p>Utilité de l'évaluation externe (UTI). Cette dimension mesure l'utilité que l'enseignant donne à l'évaluation externe afin de modifier ses pratiques d'enseignement et d'évaluation.</p> <p>1.3. <i>Les EC sont nécessaires pour compléter mes évaluations sommatives.</i></p> <p>1.10. <i>Les EC sont inutiles pour mon travail d'enseignant.</i></p> <p>1.12. <i>Je me sers des EC pour vérifier que mes pratiques d'enseignant sont adéquates.</i></p>
<p>Enjeux ressentis face à l'évaluation externe (ENJ). Cette dimension mesure le niveau de pression que les enseignants ressentent face à l'évaluation externe.</p> <p>1.2. <i>Les résultats de mes élèves aux EC me préoccupent beaucoup.</i></p> <p>1.6. <i>Les EC mettent une grosse pression sur les épaules des enseignants.</i></p> <p>1.9. <i>La réussite de mes élèves aux EC ne me stresse pas.</i></p>
<p>Qualité de l'évaluation externe (QUA). Cette dimension mesure la qualité que l'enseignant donne à l'évaluation externe dans le but d'évaluer les apprentissages des élèves.</p> <p>1.1. <i>Les exercices des EC sont d'excellente qualité.</i></p> <p>1.4. <i>Les EC sont un exemple à suivre pour l'évaluation des apprentissages des élèves.</i></p> <p>1.8. <i>Les EC n'évaluent pas les vraies capacités de mes élèves.</i></p>
<p>Concordance des pratiques de l'évaluation interne et externe (CON). Cette dimension mesure la concordance entre les pratiques évaluatives de l'enseignant (interne) et celles de l'épreuve externe.</p> <p>1.5. <i>Les EC évaluent d'autres compétences que mes évaluations sommatives.</i></p> <p>1.7. <i>Les exercices des EC ressemblent fortement à ceux de mes propres évaluations sommatives.</i></p> <p>1.11. <i>Les résultats de mes élèves aux EC sont les mêmes que dans mes évaluations sommatives.</i></p>

b) *Dimensions validées*

L'analyse factorielle exploratoire réalisée sur cette partie du questionnaire a fait ressortir 4 facteurs qui couvrent 61% de la variance observée (KMO =.71 ; test de Bartlett < 1%). Toutefois, seules les deux

premières dimensions attendue *a priori* se retrouvent dans la solution factorielle, le troisième et le quatrième facteur identifiés mélangeant les items portant sur les deux dernières dimensions (QUA et CON). Ces dernières ont donc été supprimées. L'analyse factorielle confirmatoire, dont les indices d'ajustements se sont révélés satisfaisants, a permis de confirmer l'adéquation de ce modèle à 2 dimensions aux données (NNFI =.89 ; CFI =.91 ; RMSEA =.90 ; $\chi^2/ dl = 3.78$).

Les analyses de fiabilité (alpha de Cronbach) n'ont pas non plus permis de construire les deux dernières dimensions, les valeurs étant bien inférieures au seuil requis. Les items suivants (1.1. ; 1.5. ; 1.7. ; 1.8. ; 1.11.) n'ont donc pas permis de créer les dimensions escomptées (QUA et CON). Ces dernières avaient certainement un sens trop proche l'une de l'autre et les items pas assez clairement formulés. Pour pouvoir tout de même tester ces deux dimensions, un seul item a été retenu pour chacune. L'item dont la formulation est la plus proche du nom de la dimension a été retenu. La qualité psychométrique en est donc affectée. Ce biais méthodologique est pris en compte dans l'analyse des résultats. Le tableau ci-dessous présente, pour chaque dimension, les items retenus suite à l'analyse factorielle exploratoire (les indices de saturation sont indiqués) et aux analyses de consistance interne (alpha de Cronbach).

TABLEAU 13 – Dimensions et items de la partie 1 (validée)

<i>Utilité de l'évaluation externe pour ses propres pratiques (UTI)</i>	
1.3. <i>Les EC sont nécessaires pour compléter mes évaluations sommatives.</i>	.72
1.4. <i>Les EC sont un exemple à suivre pour l'évaluation des apprentissages des élèves.</i>	.70
1.10. <i>Les EC sont inutiles pour mon travail d'enseignant.</i>	.73
1.12. <i>Je me sers des EC pour vérifier que mes pratiques d'enseignant sont adéquates.</i>	.60
Remarque : L'item 1.4. s'est greffé à cette dimension lors de l'analyse factorielle alors qu'a priori il était destiné à la dimension « Qualité perçue de l'évaluation externe ».	
Cette dimension formée par ces 4 items présente une consistance interne satisfaisante ($\alpha = .69$).	
<i>Enjeux ressentis face à l'évaluation externe (ENJ)</i>	
1.2. <i>Les résultats de mes élèves aux EC sont me préoccupent beaucoup.</i>	.82
1.6. <i>Les EC mettent une grosse pression sur les épaules des enseignants.</i>	.75
1.9. <i>La réussite de mes élèves aux EC ne me stresse pas.</i>	.83
Cette dimension formée par ces 3 items présente une consistance interne très satisfaisante ($\alpha = .77$).	
<i>Qualité de l'évaluation externe (QUA)</i>	
1.1. <i>Les exercices des EC sont d'excellente qualité.</i>	

Concordance des pratiques de l'évaluation interne et externe (CON).

1.7. *Les exercices des EC ressemblent fortement à ceux de mes propres évaluations sommatives.*

19.4.2. Partie 2 - Effets de l'évaluation externe sur les pratiques

Cette partie du questionnaire a pour objectif de mesurer les effets de l'évaluation externe sur les pratiques des enseignants du degré primaire. Les dimensions représentent les effets qui ont été identifiés par la recherche exploratoire. L'enseignant répond en faisant référence à une seule épreuve externe (comme pour la partie 1).

Nous présentons premièrement les dimensions prévues *a priori* dans notre questionnaire, puis les dimensions validées par les analyses et utilisées dans l'analyse des données.

a) Dimensions du questionnaire (a priori)

A priori (avant les analyses de validation), cette partie 2 était composée de 22 items qui investiguent 5 dimensions (tableau 14).

TABLEAU 14 – Items et dimensions de la partie 2 (*a priori*)

Harmonisation des pratiques d'évaluation sommative (HAR). Cette dimension mesure dans quelle proportion l'EE fait évoluer les pratiques d'évaluation sommative de l'enseignant.

- 2.1. *Les EC me donnent régulièrement de nouvelles idées pour évaluer mes élèves.*
- 2.5. *Sans l'exemple des EC, je n'évaluerais pas certaines compétences de mes élèves (par ex : compréhension orale).*
- 2.9. *Je prends les EC comme exemples pour créer mes propres évaluations sommatives.*
- 2.14. *Les EC n'influencent pas ma manière d'évaluer les apprentissages de mes élèves.*
- 2.16. *Les EC font évoluer mes pratiques d'évaluation sommative.*

Teaching to the test (TTT). Cette dimension mesure dans quelle proportion l'EE amène l'enseignant à utiliser des méthodes d'enseignement spécifiques à la réussite des élèves à l'épreuve (périodes de révisions, drill, épreuves de préparation).

- 2.3. *Je ne fais pas d'exercices spécifiques pour préparer mes élèves aux EC.*
- 2.8. *J'organise d'importantes périodes de révisions pour préparer mes élèves aux EC.*

2.11.	<i>Je consacre une grande partie de mon temps d'enseignement à des exercices proches de ceux des EC pour préparer mes élèves.</i>
2.19.	<i>Il est indispensable pour préparer les élèves de leur faire passer des EC des années précédentes.</i>
2.20.	<i>En vue des EC, je fais davantage de « drill » des contenus qui seront évalués.</i>
<p>Rétrécissement du curriculum enseigné (RET). Cette dimension mesure dans quelle proportion l'EE amène l'enseignant à privilégier les contenus évalués au détriment des autres.</p>	
2.6.	<i>Dans ma classe, les branches évaluées aux EC n'ont pas la priorité sur les autres.</i>
2.10.	<i>Dans mes planifications, je privilégie les contenus qui seront évalués dans les EC au détriment d'autres branches.</i>
2.12.	<i>En vue des EC, je dois mener mon programme au pas de charge et n'ai pas de temps pour certains contenus.</i>
2.17.	<i>En raison des EC, je dois renoncer aux activités hors programme (par ex : camps, sorties, ...).</i>
2.21.	<i>Certaines branches disparaissent en partie de ma grille horaire en vue des EC.</i>
<p>Développement du professionnalisme (PRO). Cette dimension mesure dans quelle proportion l'EE rend les enseignants davantage attentifs à leurs pratiques (enseignement, évaluation).</p>	
2.2.	<i>En vue des EC, je planifie mes leçons avec beaucoup plus de rigueur que d'habitude.</i>
2.4.	<i>En vue de l'EC, j'observe encore plus les apprentissages de mes élèves.</i>
2.13.	<i>Je suis davantage attentif à la qualité de mon enseignement en vue des EC.</i>
2.15.	<i>Les EC m'amènent à réfléchir davantage à mes pratiques d'enseignant.</i>
<p>Encouragement à la collaboration (COL). Cette dimension mesure dans quelle proportion l'EE amène les enseignants à collaborer davantage.</p>	
2.7.	<i>Les EC m'incitent à collaborer davantage avec des collègues.</i>
2.18.	<i>Les EC stimulent les discussions entre enseignants.</i>
2.22.	<i>J'échange davantage avec mes collègues en vue des EC.</i>

b) *Dimensions validées*

L'analyse factorielle exploratoire réalisée sur cette partie du questionnaire a tout d'abord fait apparaître 4 facteurs, les items traitant de deux effets de l'EE étant confondus (TTT et RET). Ces deux effets étant clairement définis dans la littérature et dans la recherche exploratoire, une nouvelle analyse factorielle a été menée en définissant *a priori* le nombre de 5 facteurs. Cette deuxième analyse fait ressortir 5 facteurs attendus. Ils couvrent 62% de la variance observée (KMO =.89 ; test de Bartlett < .01). L'analyse confirmatoire, dont les indices d'ajustements se sont révélés satisfaisants, a permis de

confirmer l'adéquation de ce modèle à 5 dimensions aux données (NNFI =.79 ; CFI =.84 ; RMSEA =.10 ; χ^2 / dl = 3.81).

Le tableau ci-dessous présente, pour chaque effet, les items retenus suite à l'analyse factorielle exploratoire et aux analyses de consistance interne (alpha de Cronbach).

TABLEAU 15 – Dimensions et items de la partie 2 (validée)

Harmonisation des pratiques d'évaluation sommative (HAR)		
2.1.	<i>Les EC me donnent régulièrement de nouvelles idées pour évaluer mes élèves.</i>	.80
2.5.	<i>Sans l'exemple des EC, je n'évaluerais pas certaines compétences de mes élèves (par ex : compréhension orale).</i>	.13
2.9.	<i>Je prends les EC comme exemples pour créer mes propres évaluations sommatives.</i>	.81
2.14.	<i>Les EC n'influencent pas ma manière d'évaluer les apprentissages de mes élèves.</i>	.18
2.16.	<i>Les EC font évoluer mes pratiques d'évaluation sommative.</i>	.75
Suite à l'analyse factorielle, les items 2.5. et 2.14. ont été supprimés de cette dimension.		
Après suppression de ces items, le facteur formé de ces 3 items présente une consistance interne satisfaisante ($\alpha = .73$).		
Teaching to the test (TTT)		
2.3.	<i>Je ne fais pas d'exercices spécifiques pour préparer mes élèves aux EC.</i>	.44
2.8.	<i>J'organise d'importantes périodes de révisions pour préparer mes élèves aux EC.</i>	.73
2.11.	<i>Je consacre une grande partie de mon temps d'enseignement à des exercices proches de ceux des EC pour préparer mes élèves.</i>	.73
2.19.	<i>Il est indispensable pour préparer les élèves de leur faire passer des EC des années précédentes.</i>	.77
2.20.	<i>En vue des EC, je fais davantage de « drill » des contenus qui seront évalués.</i>	.81
Ce facteur formé par ces 5 items présente une consistance interne très satisfaisante ($\alpha = .86$).		
Remarque : l'item 2.3. a été conservé malgré son indice de saturation faible. Sa suppression aurait causé une baisse de l'indice de consistance interne.		
Rétrécissement du curriculum enseigné (RET)		
2.6.	<i>Dans ma classe, les branches évaluées aux EC n'ont pas la priorité sur les autres.</i>	.43
2.10.	<i>Dans mes planifications, je privilégie les contenus qui seront évalués dans les EC au détriment d'autres branches.</i>	.70
2.12.	<i>En vue des EC, je dois mener mon programme au pas de charge et n'ai pas de temps pour certains</i>	.79

<i>contenus.</i>	
2.17. <i>En raison des EC, je dois renoncer aux activités hors programme (par ex : camps, sorties, ...).</i>	.69
2.21. <i>Certaines branches disparaissent en partie de ma grille horaire en vue des EC.</i>	.81
Ce facteur formé par ces 5 items présente une consistance interne très satisfaisante ($\alpha = .81$).	
Remarque : l'item 2.6. a été conservé malgré son indice de saturation faible. Sa suppression aurait causé une baisse de l'indice de consistance interne.	
<i>Développement du professionnalisme (PRO)</i>	
2.2. <i>En vue des EC, je planifie mes leçons avec beaucoup plus de rigueur que d'habitude.</i>	.60
2.4. <i>En vue de l'EC, j'observe encore plus les apprentissages de mes élèves.</i>	.53
2.13. <i>Je suis davantage attentif à la qualité de mon enseignement en vue des EC.</i>	.42
2.15. <i>Les EC m'amènent à réfléchir davantage à mes pratiques d'enseignant.</i>	.52
Ce facteur formé par ces 4 items présente une consistance interne satisfaisante ($\alpha = .70$).	
<i>Encouragement à la collaboration (COL)</i>	
2.7. <i>Les EC m'incitent à collaborer davantage avec des collègues.</i>	.57
2.18. <i>Les EC stimulent les discussions entre enseignants.</i>	.73
2.22. <i>J'échange davantage avec mes collègues en vue des EC.</i>	.72
Ce facteur formé par ces 3 items présente une consistance interne satisfaisante ($\alpha = .69$).	

19.4.3. Partie 3 – Conceptions de l'apprentissage

Cette partie du questionnaire a pour objectif de mesurer les conceptions qu'ont les enseignants de l'apprentissage en milieu scolaire. Pour ce faire, un outil créé et validé par Issaeva & Crahay (2010b) a été adapté pour la présente recherche. En se basant sur différentes théories de l'apprentissage et sur leurs travaux antérieurs, les auteurs ont créé une échelle qui regroupe 7 dimensions. Chaque dimension représente une possibilité de concevoir l'apprentissage. Les auteurs ont intentionnellement voulu que ces dimensions ne soient pas exclusives, l'apprentissage pouvant se voir comme multidimensionnel. Voici les 7 dimensions présentes dans cette échelle validée par ses auteurs :

- *Répétition.* Cette dimension permet de voir l'importance que les enseignants donnent à l'entraînement et la reproduction dans l'apprentissage.
- *Déséquilibre.* Cette dimension permet de mesurer si la phase de déséquilibre, dues à un obstacle ou problème à surmonter, est importante pour les enseignants.

- *Apprentissage implicite.* Ici, l'apprentissage peut être vu comme un processus qui n'est pas intentionnel et qui vient de l'expérience de l'élève à son insu.
- *Intégration de connaissances antérieures.* Cette dimension part du principe que les nouvelles connaissances se basent sur des pré-requis.
- *Eradication des anciennes connaissances.* Dans cette dimension, on conçoit que l'apprentissage se fait en remplaçant une ancienne connaissance par une nouvelle davantage « valide ».
- *Compréhension.* On part ici du principe que l'apprentissage se fait par un effort de compréhension, ce qui s'oppose à la dimension implicite.
- *Déclic.* suite à une recherche sur les difficultés de lecture, où des enseignants s'exprimaient sur l'apprentissage, les chercheurs ont intégré cette dimension qui compare l'apprentissage à un « déclic », une sorte de révélation soudaine (*insight*).

Pour atteindre nos objectifs, nous avons adapté cet outil de mesure. Nous présentons premièrement les dimensions prévues *a priori* dans notre questionnaire (et les modifications de l'échelle initiale), puis les dimensions validées par les analyses et utilisées dans l'analyse des données.

a) *Dimensions du questionnaire (a priori)*

L'échelle de Issaeva & Crahay (2010b) constitue donc un outil de mesure fiable pour la présente recherche, toutefois des adaptations ont été faites. Premièrement, les dimensions « compréhension » et « déclic » n'ont pas été intégrées dans le questionnaire, leur fondement théorique n'étant pas suffisamment justifié. De plus, les analyses statistiques présentées par les auteurs ne leur donnent, pour la première aucune existence et pour la deuxième que peu de force, la variance expliquée par cette dimension étant la plus faible (4%). Deuxièmement, dans le questionnaire de la thèse, l'échelle a été réduite à 15 items afin de raccourcir son temps de passation. 3 items ont été conservés pour chacune des 5 dimensions. Les items retenus sont ceux qui présentent les valeurs de saturation les plus importantes sur les dimensions dans les résultats de Issaeva & Crahay (2010b). Finalement, l'échelle de Likert a été modifiée pour permettre 6 modalités de réponse (4 dans l'outil initial), permettant ainsi une plus grande finesse dans la réponse.

A priori (avant les analyses de validation), cette partie 3 était composée de 15 items qui investiguent 5 dimensions (tableau 16).

TABLEAU 16 – Items et dimensions de la partie 3 (*a priori*)

<p>Répétition (REP)</p> <p>3.4. <i>La répétition n'est pas nécessaire à l'apprentissage.</i></p> <p>3.5. <i>On apprend en mettant en pratique de façon répétitive.</i></p> <p>3.9. <i>Il faut réaliser un certain nombre de fois le même type d'exercices pour apprendre.</i></p>
<p>Déséquilibre (DES)</p> <p>3.1. <i>Pour bien apprendre, il faut être placé devant un obstacle à surmonter.</i></p> <p>3.7. <i>Pour apprendre, il faut être confronté à un problème que l'on cherche à résoudre.</i></p> <p>3.12. <i>Pour développer des compétences, l'enfant doit être placé face à des situations problèmes.</i></p>
<p>Apprentissage implicite (APP)</p> <p>3.2. <i>On peut apprendre par la pratique ; pas toujours besoin de comprendre.</i></p> <p>3.11. <i>On peut apprendre sans comprendre, par l'expérience directe des choses.</i></p> <p>3.13. <i>Il n'y a pas toujours besoin de comprendre pour apprendre.</i></p>
<p>Eradication des connaissances antérieures (ERA)</p> <p>3.3. <i>Apprendre, c'est remplacer des conceptions antérieures par des conceptions nouvelles.</i></p> <p>3.8. <i>Pour apprendre, il faut d'abord abandonner ce qu'on savait auparavant.</i></p> <p>3.15. <i>Apprendre, c'est pouvoir remplacer une connaissance par une autre.</i></p>
<p>Intégration des connaissances antérieures (INT)</p> <p>3.6. <i>On pourrait dire que les prérequis sont les bases sur lesquelles on construit la suite des apprentissages.</i></p> <p>3.10. <i>Apprendre, c'est établir des liens entre de nouvelles informations et des savoirs possédés antérieurement.</i></p> <p>3.14. <i>Pour apprendre, il faut partir de ce que l'on sait déjà.</i></p>

b) Echelle validée

L'analyse factorielle exploratoire réalisée sur cette partie 3 du questionnaire a fait ressortir 5 facteurs qui représentent les 5 dimensions présentées et qui couvrent 64% de la variance observée (KMO =.684 ; test de Bartlett < .01). L'analyse confirmatoire, dont les indices d'ajustements se sont révélés satisfaisants, a permis de confirmer l'adéquation de ce modèle à 5 dimensions aux données (NNFI =.864 ; CFI =.923 ; RMSEA =.059 ; $\chi^2/ dl = 2.077$). Le tableau ci-dessous présente, pour chaque dimension, les items retenus suite à l'analyse factorielle exploratoire (les indices de saturation sont indiqués) et aux analyses de consistance interne (alpha de Cronbach).

TABLEAU 17 – Items et dimensions de la partie 3 (a priori)

Répétition (REP)		
3.4.	La répétition n'est pas nécessaire à l'apprentissage.	.60
3.5.	On apprend en mettant en pratique de façon répétitive.	.53
3.9.	Il faut réaliser un certain nombre de fois le même type d'exercices pour apprendre.	.42
Suite à l'analyse factorielle et aux analyses de consistance interne, l'item 3.4. a été supprimé de cette dimension pour avoir une consistance interne satisfaisante ($\alpha = .723$).		
Déséquilibre (DES)		
3.1.	Pour bien apprendre, il faut être placé devant un obstacle à surmonter.	.72
3.7.	Pour apprendre, il faut être confronté à un problème que l'on cherche à résoudre.	.81
3.12.	Pour développer des compétences, l'enfant doit être placé face à des situations problèmes.	.84
Ce facteur formé par ces 3 items présente une consistance interne satisfaisante ($\alpha = .75$).		
Apprentissage implicite (APP)		
3.2.	On peut apprendre par la pratique ; pas toujours besoin de comprendre.	.78
3.11.	On peut apprendre sans comprendre, par l'expérience directe des choses.	.85
3.13.	Il n'y a pas toujours besoin de comprendre pour apprendre.	.80
Ce facteur formé par ces 3 items présente une consistance interne satisfaisante ($\alpha = .76$).		
Eradication des connaissances antérieures (ERA)		
3.3.	Apprendre, c'est remplacer des conceptions antérieures par des conceptions nouvelles.	.77
3.8.	Pour apprendre, il faut d'abord abandonner ce qu'on savait auparavant.	.71
3.15.	Apprendre, c'est pouvoir remplacer une connaissance par une autre.	.78
Ce facteur formé par ces 3 items présente une consistance interne satisfaisante ($\alpha = .67$).		
Intégration des connaissances antérieures (INT)		
3.6.	On pourrait dire que les prérequis sont les bases sur lesquelles on construit la suite des apprentissages.	.57
3.10.	Apprendre, c'est établir des liens entre de nouvelles informations et des savoirs possédés antérieurement.	.73
3.14.	Pour apprendre, il faut partir de ce que l'on sait déjà.	.72
Ce facteur formé par ces 3 items présente une consistance interne peu satisfaisante ($\alpha = .57$).		

19.4.4. Partie 4 – Conceptions de l'auto-efficacité personnelle

Cette partie du questionnaire a pour but de mesurer l'efficacité que les enseignants attribuent à leurs propres pratiques pédagogiques. L'échelle de mesure utilisée pour cette recherche a été adaptée de celle créée par Issaeva & Crahay (2010b). Les auteurs se sont basés sur le champ théorique de l'auto-efficacité pédagogique de l'enseignant (*teachers efficacy beliefs*) et sur les instruments de mesure préexistants (Bandura, 2003; Dussault, Villeneuve, & Deaudelin, 2001; Tschannen-Moran & Woolfolk Hoy, 2001).

Elle permet d'appréhender deux dimensions : la perception de l'efficacité générale et la perception personnelle. Aussi, l'échelle s'organise autour de 4 domaines de compétence : enseignement ; gestion de la classe ; contrôle sur les influences sociales ; motivation des élèves. Par conséquent, l'enseignant peut estimer ses compétences selon ces 4 domaines (Issaeva & Crahay, 2010a, p. 4).

Pour atteindre nos objectifs, nous avons adapté cet outil de mesure. Nous présentons premièrement les dimensions prévues *a priori* dans notre questionnaire (et les modifications de l'échelle initiale), puis les dimensions validées par les analyses et utilisées dans l'analyse des données.

a) Dimensions du questionnaire (a priori)

Nous avons adapté l'échelle de Issaeva & Crahay (2010b) pour le questionnaire de la thèse. Premièrement, seules les dimensions ayant trait à l'« efficacité personnelle » des enseignants ont été prises en compte. En effet, au regard de l'analyse factorielle de ces auteurs, les dimensions de l'« efficacité générale » n'expliquent que peu de variance (< 5%) et investiguent des domaines qui ne semblent pas pertinent pour la présente étude. Deuxièmement, la dimension «contrôle sur les influences sociales » n'a pas été intégrée dans le questionnaire. Il semble, selon la littérature et la recherche exploratoire n'avoir aucun lien sur les effets de l'EE. Troisièmement, un domaine supplémentaire a été ajouté spécifiquement pour cette recherche. Il s'agit de l'évaluation des apprentissages des élèves. Il semblait important d'intégrer ce domaine de pratiques, étant identifié par la recherche exploratoire comme un facteur d'influence sur les effets de l'EE. Finalement, l'échelle de Likert a été modifiée pour permettre 6 modalités de réponse, permettant ainsi une plus grande finesse dans la réponse.

Avant les analyses de validation, l'échelle était donc composée de 15 items qui investiguent 5 dimensions (tableau 18).

TABLEAU 18 – Items et dimensions de la partie 4 (*a priori*)

<p>Maîtrise des savoirs (MSA). L'enseignant estime son efficacité personnelle à amener ses élèves à la maîtrise des apprentissages.</p>	
4.3.	<i>Lorsque les élèves arrivent à maîtriser aisément un nouveau concept, c'est parce que j'ai trouvé des moyens efficaces pour enseigner.</i>
4.7.	<i>Je me sens capable d'amener tous les élèves à réussir une tâche même lorsqu'elle est difficile.</i>
4.12.	<i>Je me sens capable d'amener tous les élèves à la maîtrise des savoirs enseignés à l'école.</i>
<p>Gestion de la classe (GES). L'enseignant estime son efficacité personnelle à gérer les problèmes de comportement des élèves en classe.</p>	
4.1.	<i>Je me sens capable de remédier efficacement aux problèmes de discipline en classe.</i>
4.8.	<i>Il m'est difficile de gérer les problèmes de comportement en classe.</i>
4.10.	<i>Si un élève de ma classe est bruyant et dérange, je me sens capable de lui faire respecter les règles de la classe.</i>
<p>Motivation des élèves (MOT). L'enseignant estime son efficacité personnelle à motiver ses élèves pour les activités scolaires.</p>	
4.2.	<i>Je me sens capable de motiver les élèves qui montrent peu d'intérêt pour les tâches scolaires.</i>
4.5.	<i>Je me sens capable d'intéresser tous les élèves aux activités que j'organise.</i>
4.9.	<i>Même avec un bon enseignant, il y a toujours des élèves non motivés.</i>
<p>Evaluation des apprentissages (EVA). L'enseignant estime son efficacité personnelle à pouvoir évaluer les apprentissages de ses élèves.</p>	
4.4.	<i>Il m'est difficile d'évaluer les apprentissages de mes élèves avec une grande précision.</i>
4.6.	<i>Après une séquence d'enseignement, je me sens capable de repérer les difficultés et les réussites de mes élèves.</i>
4.11.	<i>Je me sens capable de déterminer de manière précise si un élève maîtrise ou non une matière.</i>

b) Dimensions validées

L'analyse factorielle exploratoire réalisée sur cette la partie 4 du questionnaire a fait ressortir 4 facteurs qui couvrent 60% de la variance observée (KMO =.77 ; test de Bartlett < .01). Les trois premiers facteurs représentent chacun une dimension présentée. Par contre, le quatrième facteur lie sans logique quelques items. La dimension « Motivation des élèves » n'est pas reconnue par l'analyse factorielle et a donc été supprimée. L'analyse factorielle confirmatoire, dont les indices d'ajustements se sont révélés satisfaisants, a permis de confirmer l'adéquation de ce modèle à 3 dimensions aux données (NNFI =.95 ; CFI =.98 ; RMSEA =.04 ; $\chi^2/ dl = 1.43$).

Le tableau ci-dessous présente, pour chaque dimension, les items retenus suite à l'analyse factorielle exploratoire (les indices de saturation sont indiqués) et aux analyses de consistance interne (alpha de Cronbach).

TABLEAU 19 – Items et dimensions de la partie 4 (validée)

Maîtrise des savoirs (MSA). L'enseignant estime son efficacité personnelle à amener ses élèves à la maîtrise des apprentissages.		
4.2.	Je me sens capable de motiver les élèves qui montrent peu d'intérêt pour les tâches scolaires.	.65
4.3.	Lorsque les élèves arrivent à maîtriser aisément un nouveau concept, c'est parce que j'ai trouvé des moyens efficaces pour enseigner.	.41
4.5.	Je me sens capable d'intéresser tous les élèves aux activités que j'organise.	.72
4.7.	Je me sens capable d'amener tous les élèves à réussir une tâche même lorsqu'elle est difficile.	.73
4.12.	Je me sens capable d'amener tous les élèves à la maîtrise des savoirs enseignés à l'école.	.74
Suite à l'analyse factorielle et aux analyses de consistance interne, les items 4.2. et 4.3. ont été supprimés de cette dimension pour avoir une consistance interne satisfaisante ($\alpha = .72$).		
Remarque : Les items 4.2. et 4.5. se sont greffés à cette dimension lors de l'analyse factorielle exploratoire, alors qu'a priori ils étaient destinés à la dimension « motivation des élèves ».		
Gestion de la classe (GES)		
4.1.	Je me sens capable de remédier efficacement aux problèmes de discipline en classe.	.82
4.8.	Il m'est difficile de gérer les problèmes de comportement en classe.	.71
4.10.	Si un élève de ma classe est bruyant et dérange, je me sens capable de lui faire respecter les règles de la classe.	.76
Ce facteur formé par ces 3 items présente une consistance interne satisfaisante ($\alpha = .71$).		
Evaluation des apprentissages (EVA)		
4.4.	Il m'est difficile d'évaluer les apprentissages de mes élèves avec une grande précision.	.08
4.6.	Après une séquence d'enseignement, je me sens capable de repérer les difficultés et les réussites de mes élèves.	.78
4.11.	Je me sens capable de déterminer de manière précise si un élève maîtrise ou non une matière.	.76
Suite à l'analyse factorielle et aux analyses de consistance interne, l'item 4.4. a été supprimé de cette dimension pour avoir une consistance interne satisfaisante ($\alpha = .68$).		

19.4.5. Partie 5 - Conceptions de l'évaluation scolaire

Cette partie du questionnaire a pour but de mesurer les conceptions des enseignants à propos de l'évaluation des apprentissages des élèves. L'échelle de mesure utilisée pour cette recherche a été adaptée de celle créée par Issaeva & Crahay (2010b). Elle a été construite « [...] à partir du présupposé que l'évaluation scolaire peut être perçue sous deux angles : soit comme un processus formateur visant la maîtrise des savoirs, soit comme un processus normatif valorisant la performance » (p.40). Le questionnaire validé par les auteurs regroupe 5 facteurs qui se rapportent à l'une des deux visions de l'évaluation, des corrélations positives étant observées entre les facteurs ayant la même visée évaluative :

- a) *Normative*. 3 facteurs se recoupent sur une vision plutôt normative de l'évaluation (évaluation vu comme un moyen de repérer les élèves en fonction de leurs aptitudes scolaires, l'évaluation vu comme une sanction et l'évaluation vue comme un bilan de compétences) ($r = .44$ et $.54$)
- b) *Formative*. 2 facteurs regroupent les items qui vont dans le sens d'une vision formative de l'évaluation (repérage des élèves en difficulté et de leurs progrès et une fonction régulatrice) ($r = .44$).

Pour atteindre nos objectifs, nous avons adapté cet outil de mesure. Nous présentons premièrement les dimensions prévues *a priori* dans notre questionnaire (et les modifications de l'échelle initiale), puis les dimensions validées par les analyses et utilisées dans l'analyse des données.

a) Dimensions du questionnaire (*a priori*)

Pour cette recherche, l'échelle proposée par Issaeva & Crahay (2010a) a été reprise avec quelques modifications. Elle se compose des 2 dimensions : normative et formative. Ces deux visions ne sont pas exclusives, les enseignants devant se positionner par rapport à chacune d'elle. Pour que la passation du questionnaire soit plus brève, le nombre d'items a été réduit (8 au lieu de 20). Les items sont regroupés en deux dimensions, mais ont été sélectionnés dans chacun des 5 facteurs identifiés par Issaeva & Crahay (2010a). Le choix des items s'est fait selon l'importance de l'indice de saturation. Aussi, les items ont été modifiés, on ne parle plus spécifiquement de français ou de mathématiques mais de scolarité en général. Enfin, l'échelle de Likert a été modifiée pour permettre 6 modalités de réponse, permettant ainsi une plus grande finesse dans la réponse.

Avant les analyses de validation, l'échelle était donc composée de 8 items qui investiguent 2 dimensions (tableau 20).

TABLEAU 20 – Items et dimensions de la partie 5 (*a priori*)

<i>Evaluation formative (EVALFORM)</i>	
5.1.	<i>L'évaluation permet de cerner les progrès que les élèves ont réalisés.</i>
5.3.	<i>Les évaluations permettent de repérer les élèves qui ont besoin d'une explication supplémentaire.</i>
5.6.	<i>L'analyse des résultats des élèves aux évaluations permet d'ajuster l'enseignement en fonction de leurs faiblesses.</i>
5.8.	<i>Les évaluations permettent de repérer ce qu'il faut retravailler avec mes élèves.</i>
<i>Evaluation normative (EVALNORM)</i>	
5.2.	<i>Les résultats des évaluations permettent d'identifier les élèves qui n'ont pas d'aptitudes scolaires.</i>
5.4.	<i>Les évaluations sanctionnent les élèves qui n'ont pas travaillé.</i>
5.5.	<i>Les évaluations permettent de repérer les bons et les mauvais élèves.</i>
5.7.	<i>Les évaluations sont l'occasion pour les élèves de montrer ce qu'ils valent.</i>

b) Dimensions validées

L'analyse factorielle exploratoire réalisée sur la partie 5 du questionnaire a fait ressortir 2 facteurs qui couvrent 53% de la variance observée (KMO =.730 ; test de Bartlett < .01). Ces facteurs représentent les 2 dimensions présentées. L'analyse factorielle confirmatoire, dont les indices d'ajustements se sont révélés satisfaisants, a permis de confirmer l'adéquation de ce modèle à 2 dimensions aux données (NNFI =.95 ; CFI =.98 ; RMSEA =.05 ; $\chi^2/ dl = 1.75$).

Les tableaux ci-dessous présentent, pour chaque dimension, les items retenus suite à l'analyse factorielle exploratoire et aux analyses de consistance interne.

TABLEAU 21 – Items et dimensions de la partie 5 (*validée*)

<i>Evaluation formative (EVALFORM)</i>	
5.1.	<i>L'évaluation permet de cerner les progrès que les élèves ont réalisés.</i> .58
5.3.	<i>Les évaluations permettent de repérer les élèves qui ont besoin d'une explication supplémentaire.</i> .75
5.6.	<i>L'analyse des résultats des élèves aux évaluations permet d'ajuster l'enseignement en fonction de leurs faiblesses.</i> .76
5.8.	<i>Les évaluations permettent de repérer ce qu'il faut retravailler avec mes élèves.</i> .84
Suite à l'analyse factorielle et aux analyses de consistance interne, l'item 5.1. a été supprimé de cette dimension pour avoir une consistance interne satisfaisante ($\alpha =.73$).	

Evaluation normative (EVALNORM)

5.2.	<i>Les résultats des évaluations permettent d'identifier les élèves qui n'ont pas d'aptitudes scolaires.</i>	.69
5.4.	<i>Les évaluations sanctionnent les élèves qui n'ont pas travaillé.</i>	.61
5.5.	<i>Les évaluations permettent de repérer les bons et les mauvais élèves.</i>	.82
5.7.	<i>Les évaluations sont l'occasion pour les élèves de montrer ce qu'ils valent.</i>	.64

Ce facteur formé par ces 4 items présente une consistance interne satisfaisante ($\alpha = .65$).

19.4.6. Partie 6 - Informations individuelles et professionnelles

Cette partie du questionnaire a pour objectif de connaître le profil des enseignants qui ont répondu au questionnaire. Il contient des informations personnelles (âge, sexe) mais surtout des données sur le profil professionnel (formation initiale et continue, degré d'enseignement, expérience, type d'école, lieu d'enseignement, type et degré de collaboration).

19.5. Traitement des données

Les données issues du questionnaire ont été codées. Elles ont pu être ensuite analysées à l'aide de différents logiciels de statistiques SPSS et AMOS pour mener les différents tests.

20. Présentation des résultats de la seconde partie empirique

Les résultats issus du questionnaire sont présentés séparément pour chaque type d'évaluation externe :

- *EE de référence* (épreuves de références des cycles 1-2P et 3-4P)⁸³ ;
- *EE d'orientation* (épreuve cantonale PPO du cycle 5-6P).

Nous décrivons tout d'abord les effets de l'EE identifiés sur les pratiques des enseignants. Puis, nous décrivons les perceptions de l'EE et croyances des enseignants à propos de l'enseignement. Enfin, la présentation des résultats est conclue par une modélisation des liens entre ces variables.

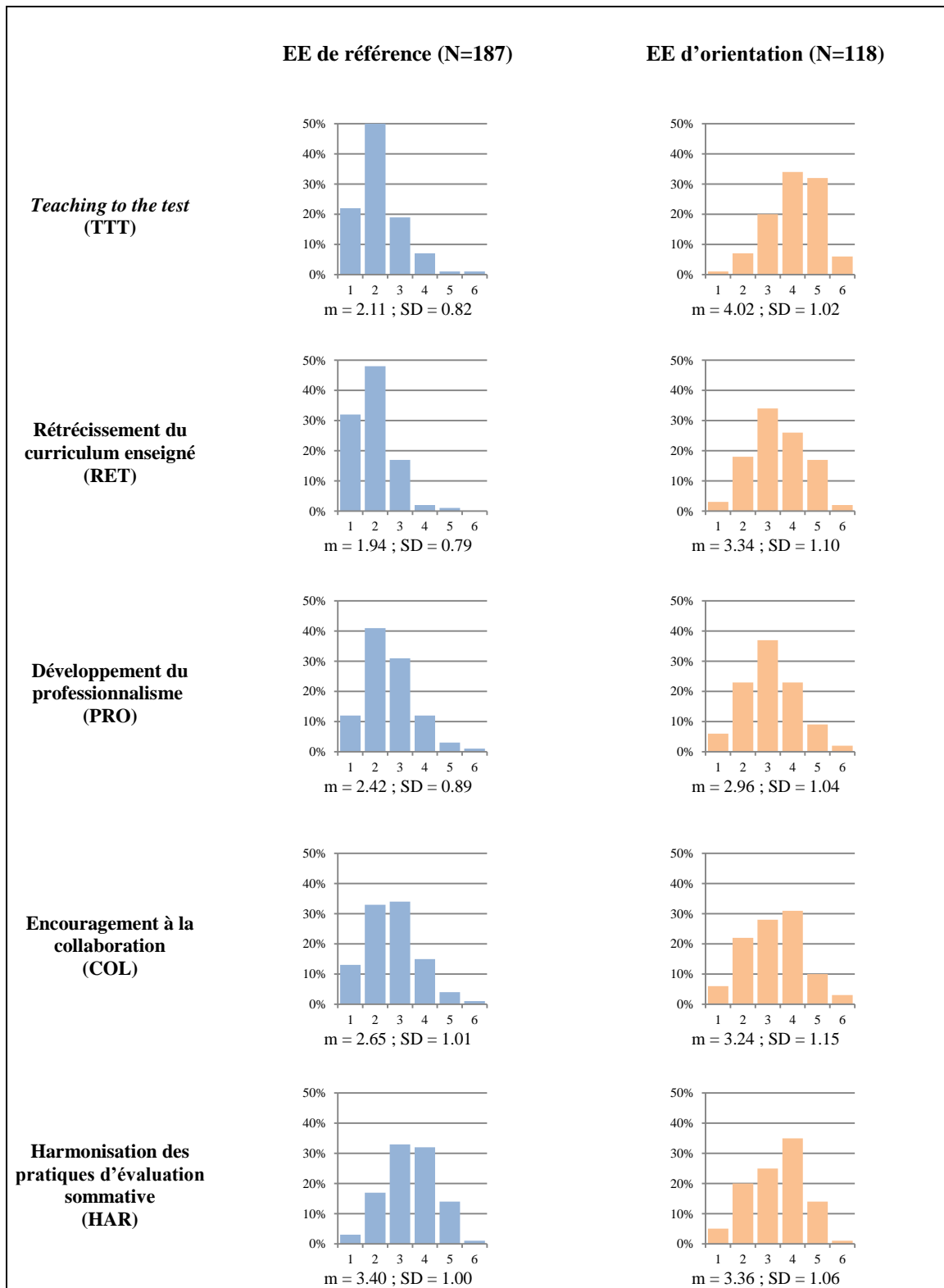
Les résultats présentés dans ce point sont interprétés conjointement dans le chapitre 5 avec ceux issus de la première partie exploratoire.

20.1. Effets de l'évaluation externe

Le tableau 22 présente la description des cinq effets qui ont été identifiés dans la partie exploratoire de la thèse (chapitre 3). Les graphiques représentent la fréquence des réponses (indiquée par le pourcentage de répondants) relatives aux échelles de Likert (1 « pas du tout d'accord » à 6 « tout à fait d'accord »).

⁸³ Les analyses démontrent que les données pour les épreuves de référence du cycle 1-2P et celles de 3-4P ne diffèrent pas significativement, elles sont donc agrégées dans un même groupe. La précaution a été prise de vérifier ce constat pour chaque analyse.

TABLEAU 22 – Description des effets de l’EE



Tout d’abord, la bonne distribution de ces variables est à noter. Elles ne s’éloignent que modérément de la loi de distribution normale, leur asymétrie (Skewness) étant inférieure à 1 (en valeur absolue) et leur aplatissement (Kurtosis) étant inférieur à 1,5 (en valeur absolue). Une seule exception est observée, il s’agit de l’effet de « rétrécissement du curriculum » pour l’épreuve de référence qui présente un léger effet plancher (Skewness = 1.45 ; Kurtosis = 4.07). Il est à rappeler que les items du questionnaire ont été formulés dans le but d’observer surtout la variance des réponses des enseignants, les réponses s’approchant de 5 et 6 représentant des effets forts, ceux de 3 ou 4 des effets moyens et 1 et 2 des effets faibles. La bonne variabilité (les résultats sont plutôt bien répartis sur l’échelle de Likert) de ces effets représente un élément essentiel pour cette étude. A l’évidence, ces effets ne sont pas perçus de manière uniforme par les enseignants et/ou pour toutes les épreuves, ce qui va permettre de répondre aux questions de recherche et hypothèses. La figure 22 présente la moyenne des scores pour chaque effet, selon le type d’EE.

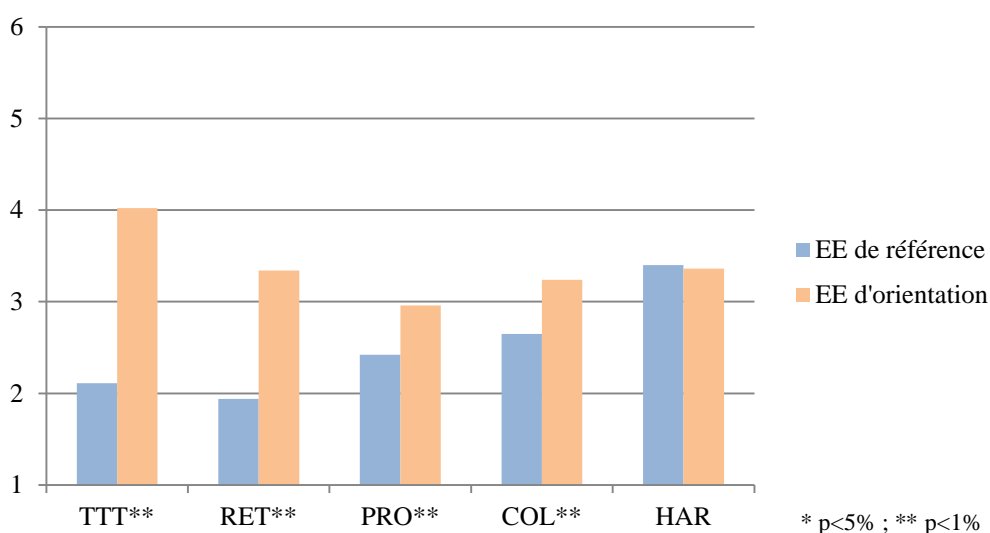


FIGURE 22 – Moyennes des effets de l’évaluation externe sur les pratiques des enseignants selon le type d’épreuve

La figure 22 permet d’observer les différences entre type d’EE. L’EE d’orientation amène très clairement les enseignants à davantage de préparation spécifique au test (TTT ; $t_{(301)}=17.09$; $p<1\%$) et de rétrécissement du curriculum (RET ; $t_{(301)}=12.03$; $p<1\%$) que les épreuves de référence ; elles « responsabilisent » également davantage les enseignants (RET ; $t_{(301)}=4.85$; $p<1\%$) et les amènent davantage à collaborer (COL ; $t_{(301)}=4.71$; $p<1\%$). Toutes ces différences sont statistiquement significatives ($p<.01$). Par contre, aucune différence statistique n’est observée au sujet de l’harmonisation des pratiques d’évaluation sommative (HAR ; $t_{(301)}=0.35$; NS).

Des corrélations sont à relever entre ces différents effets (tableau 23).

TABLEAU 23 – Corrélations entre effets de l’EE

	TTT	RET	PRO	COL	HAR
TTT		.54**	.45**	.35**	.09
RET	.69**		.36**	.24**	.09
PRO	.43**	.44**		.46**	.44**
COL	.29**	.36**	.54**		.28**
HAR	.03	-.09	.25**	.28**	

EE d’orientation (N=118)

EE de référence (N=187)

* P<5% ; ** P<1%

Nous remarquons que tous les effets corrént avec les autres et de manière plutôt importante⁸⁴(en foncé), à l’exception de l’effet d’harmonisation des pratiques d’évaluation (HAR) qui ne corréle ni avec TTT et ni avec RET. Ces corrélations sont toutes positives et significatives (p < .01), ce qui démontre que plus un enseignant perçoit un effet, plus il perçoit les autres. Ces corrélations sont semblables pour les deux types d’épreuves, mais généralement plus fortes pour l’EE d’orientation.

Les corrélations les plus fortes sont, pour les deux types d’épreuves, celles observées entre TTT et RET. Ces deux variables étaient déjà confondues lors de l’analyse factorielle exploratoire (ACP). Il est à remarquer que la variable PRO corréle de manière importante avec toutes les autres (r > .40).

⁸⁴ Corrélacion : autour de .10 = faible ; autour de .30 = modérée ; plus de .50 = forte (Cohen, 1988)

20.2. Profil personnel et professionnel des enseignants

Les résultats présentés précédemment démontrent l'existence des effets de l'EE sur les pratiques des enseignants dans le contexte de la présente recherche. Les analyses ont également démontré des différences d'effets importantes dues aux types de dispositifs externes mis en place (de référence vs d'orientation). Toutefois, pour chacun de ces dispositifs subsistent des différences importantes entre les enseignants et cela pour chaque type d'effet. L'un des objectifs de la présente étude est justement d'expliquer cette variance pour chaque type d'épreuve externe. Le présent point fait état des analyses qui ont été réalisées dans ce but en explorant spécifiquement l'influence du profil individuel et professionnel des enseignants.

Toutes les variables individuelles et professionnelles qu'investiguait la partie 6 du questionnaire ont été testées (voir p.156). Toutefois, les résultats démontrent que, parmi celles-ci, seule l'expérience professionnelle influence certains effets.

20.2.1. Lien entre l'expérience professionnelle et effets de l'EE

Afin de mesurer l'expérience professionnelle des enseignants, des catégories ont été proposées dans le questionnaire basées sur la durée d'engagement dans la profession (0-2 ans, 3-4 ans, 5-10 ans, 11-20 ans, 21-30 ans, + de 30 ans)⁸⁵, ceci pour faciliter les réponses. La figure 23 montre la répartition des participants quant à leur expérience professionnelle⁸⁶. Cette répartition nous permet de comparer ces groupes (que nous avons affinés pour observer les effets de l'EE sur les enseignants débutants).

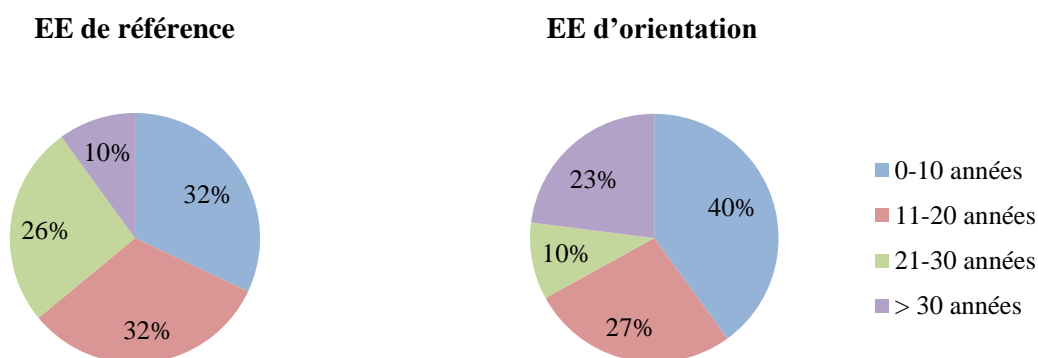


FIGURE 23 – Expérience professionnelle des participants

⁸⁵ Ces catégories ont été proposées afin de permettre au répondant de faire une estimation de son expérience (sans devoir compter exactement). Néanmoins, il aurait été préférable pour les analyses de demander le nombre d'années, sans les catégoriser a priori.

⁸⁶ Le questionnaire a également recensé l'expérience professionnelle des enseignants dans les cycles de la scolarité, Néanmoins, les deux variables ne différant pas significativement, l'expérience totale a été choisie dans les analyses.

Un seul lien significatif existe entre les années d'expérience et les effets de l'EE. Pour les épreuves de référence, plus l'enseignant a une longue carrière professionnelle, plus il constate l'effet de développement du professionnalisme ($\rho = .202$; $p < .01$).

Toutefois, les analyses montrent des différences spécifiques pour certains groupes. Pour l'EE de référence (figure 24), les enseignants les plus expérimentés (+ de 30 années d'expérience) perçoivent davantage l'effet de développement du professionnalisme ($m=3.03$) que toutes les autres catégories ($F_{(3,178)} = 4.91$; $p < 1\%$). Aussi, ce même groupe perçoit davantage l'effet d'harmonisation des pratiques d'évaluation sommative ($m=4.02$) que toutes les autres catégories ($F_{(3,178)} = 2.86$; $p < 5\%$).

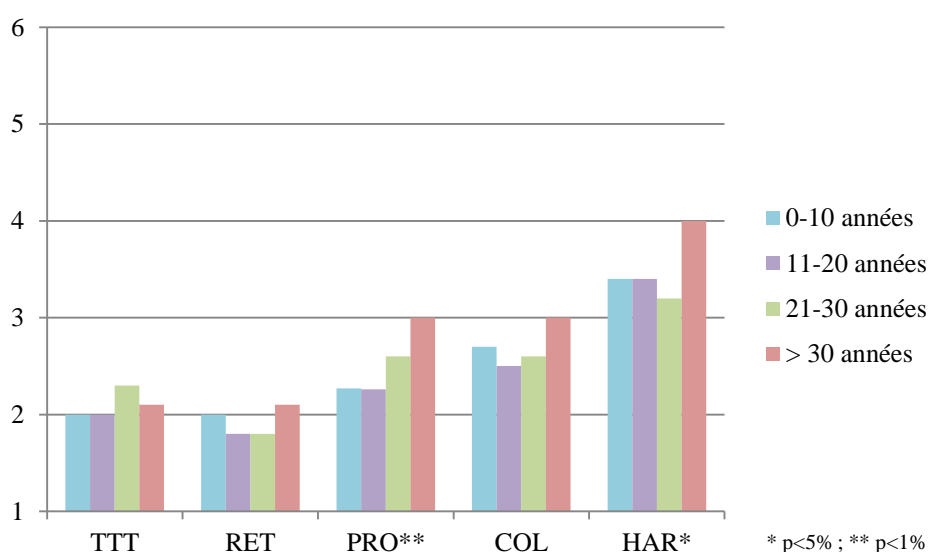


FIGURE 24 – Expérience professionnelle et effets de l'EE (EE de référence)

Egalement, si l'on regarde plus finement les catégories, on peut observer que pour les épreuves de référence, les enseignants les moins expérimentés (moins de 3 ans d'expériences) rapportent pratiquer davantage de TTT ($F_{(5,176)}=2.23$; $p = 5\%$). Pour les épreuves d'orientation, les enseignants les moins expérimentés (0-2 et 3-4 années d'expérience) ressentent davantage de développement du professionnalisme ($m=3.50$) ($F_{(5,111)}=2.65$; $p < 5\%$).

20.2.2. Autres variables personnelles et professionnelles et liens avec les effets de l'EE

D'autres variables individuelles et professionnelles ont été testées (analyses de variances simples et multiples, analyses de corrélation).

- *Genre*. Aucune différence significative n'est observée entre femmes et hommes.
- *Degré d'activité professionnelle*. Aucune différence significative n'est observée entre les enseignants à temps plein et à temps partiel.

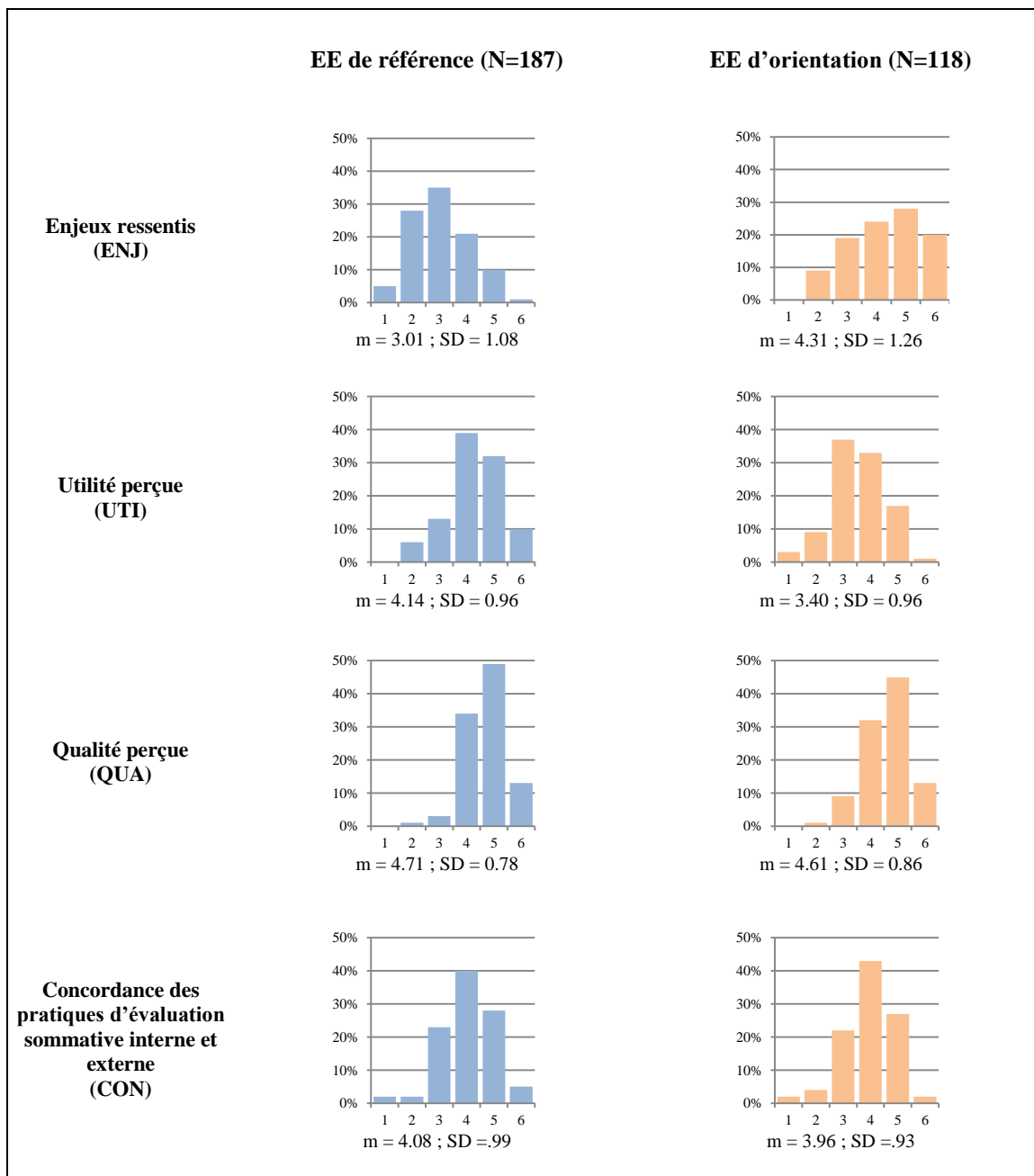
- *Parcours mixte.* Aucune différence significative n'est observée entre les enseignants qui ont toujours enseigné dans le même cycle et ceux qui ont travaillé également dans d'autres degrés.
- *Année d'évaluation externe.* Les enseignants qui ont répondu au questionnaire se trouvaient en année d'évaluation externe (2P, 4P ou 6P et double-degré) ou dans une année sans évaluation externe (1P, 3P, 5P). Les analyses montrent que ce facteur temporel de la passation n'influence pas les effets qu'ils perçoivent de l'EE, aucune différence statistique n'a été observée.
- *Double degré.* Aucune différence significative n'est observée entre les enseignants des classes à degrés multiples et les autres.
- *Formation complémentaire.* Aucune différence n'est observée entre les enseignants qui ont suivis des formations complémentaires et les autres.
- *Conception d'épreuves externe.* Aucune différence significative n'est observée entre les enseignants qui ont participé à la conception d'épreuves externes ou non.
- *Type d'école.* Les effets de l'EE ne sont pas perçus différemment par les enseignants selon le type de commune où ils enseignent (ville – périurbain – campagne).
- *Âge.* L'âge n'a pas d'influence sur les effets de l'EE.

20.3. Perceptions de l'évaluation externe

20.3.1. Description des perceptions de l'EE

Le questionnaire a permis d'explorer certaines perceptions qu'ont les enseignants de l'évaluation externe, ceci pour chaque type d'évaluation externe. Le tableau 24 en fait la description. Les graphiques représentent la fréquence des réponses (indiquée par le pourcentage de répondants) relatives aux échelles de Likert (1 « pas du tout d'accord » à 6 « tout à fait d'accord »).

TABLEAU 24 – Description des perceptions de l'EE



Pour chaque type d'EE, ces quatre variables forment des distributions normales (les valeurs de Skewness et Kurtosis sont satisfaisantes). Il est à rappeler que les dimensions QUA et CON, n'ayant pas montré une validité suffisante lors des analyses factorielles confirmatoires, ne sont calculées que sur un seul item. De plus, il est important de noter que les enseignants ont des perceptions très diverses d'une même épreuve externe. Il serait alors erroné de conclure que les enseignants ont une perception uniforme de chaque type d'épreuve externe. Comme le montre les distributions (tableau 24), ils ne sont par exemple que la moitié à ressentir de forts enjeux face à l'épreuve d'orientation.

La figure 25 présente pour chacun des effets, les différences statistiques entre les divers types de dispositifs (*t* de Student).

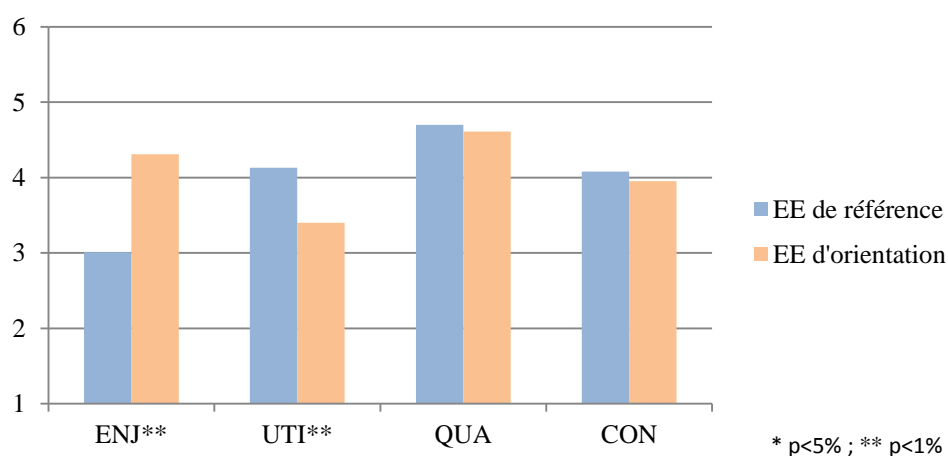


FIGURE 25 – Les perceptions de l'EE par les enseignants selon le type d'épreuve externe

Comme démontré par la figure 25, les enseignants ressentent davantage d'enjeux pour l'évaluation externe d'orientation (ENJ ; $t_{(301)}=9.58$; $p<1\%$). Au contraire, les enseignants accordent davantage d'utilité pour leurs pratiques à l'évaluation externe de référence (UTI ; $t_{(301)}=6.49$; $p<1\%$). Ces différences sont statistiquement significatives. Aucune différence statistique n'est observée entre les différents types d'épreuve externe, sur la qualité de l'EE (QUA ; $t_{(293)}=2.87$; NS) et la concordance des pratiques évaluatives internes et externes (PRO ; $t_{(296)}=1.04$; NS).

Plusieurs corrélations ont été observées entre effets de l'EE (tableau 25).

TABLEAU 25 – Corrélation entre perceptions de l’EE

	ENJ	UTI	QUA	CON
ENJ		.12	-.04	-.05
UTI	-.05		.40**	.22**
QUA	-.08	.20*		.23**
CON	-.01	.28**	.18	

EE d’orientation (N=118)

EE de référence (N=187)

* P<5% ; ** P<1%

Pour les deux types d’EE la variable UTI corrèle avec les variables QUA et CON. Pour les épreuves de référence, les variables QUA et CON corrèlent. Il est à noter qu’il n’existe pas de corrélation entre les enjeux et l’utilité, ces deux variables ne dépendent donc pas l’une de l’autre.

Quelques différences statistiques dues au profil des enseignants ont pu être observées au sujet des perceptions de l’EE.

- *Enjeux.* Pour les épreuves d’orientation, plus leur pourcentage de travail est élevé, plus les enseignants perçoivent d’enjeux ($r = .237$; $p < .01$). Les enseignants les moins expérimentés (moins de 10 ans d’expérience) ($m = 4.69$) ressentent également davantage d’enjeux que les autres ($m = 4.11$) ($F_{(1,114)} = 3.65$; $p < 5\%$).
- *Utilité.* Dans le cadre des épreuves de référence, les enseignants qui sont en charge d’une classe à double degré ($m = 3.85$) perçoivent moins d’utilité aux EE que les autres ($m = 4.23$) ($F_{(1,114)} = 5.65$; $p = .019$). Au contraire, dans le cadre des épreuves d’orientation, les enseignants de classe à double degrés ($m = 3.88$) leurs donnent davantage d’utilité ($m = 3.29$) ($F_{(1,180)} = 8.66$; $p < 1\%$).
- *Qualité.* Les enseignants les plus expérimentés (> 30 ans) ($m = 5.92$) trouvent les épreuves de références de meilleures qualité que les autres ($m = 4.61$) ($F_{(1, 114)} = 4.03$; $p = .008$).

20.3.2. Liens entre effets de l'EE et leurs perceptions

Suivant le constat énoncé précédemment, la présente recherche se propose de comparer les effets de l'évaluation externe non plus seulement par une opposition de deux types figés d'évaluation externe, mais aussi par une vision plus individualisée, les enseignants n'ayant pas des perceptions uniformes de chacun des dispositifs. Le tableau 26 met en évidence les corrélations entre les effets et les conceptions pour les deux types d'épreuve externe (pour l'EE de référence dans les colonnes de gauche, pour l'EE d'orientation dans celles de droite).

TABLEAU 26 – Corrélations entre perceptions et effets de l'évaluation externe

	TTT		RET		PRO		COL		HAR	
	référence	orientation	référence	orientation	référence	orientation	référence	orientation	référence	orientation
ENJ	.29**	.47**	.25**	.51**	.27**	.25**	.17**	.13	.09	.00
UTI	.13	.06	.07	-.05	.21**	.16	.14	.13	.45**	.49**
QUA	-.02	.21*	-.09	-.03	.09	.01	.03	.02	.25**	.15
CON	-.06	.04	-.08	-.07	-.04	-.05	-.02	-.05	.18*	.13

* p<5% ; ** p<1%

Au regard du tableau 26, il est possible de conclure que certaines perceptions de l'évaluation externe par les enseignants (ENJ et UTI) influencent ses effets. Ainsi, pour les deux types d'épreuve, plus les enjeux sont jugés comme importants, plus les effets « néfastes » (*teaching to the test*, rétrécissement du curriculum) augmentent. Aussi, plus les enseignants perçoivent l'épreuve comme utile pour leurs pratiques, plus l'effet d'harmonisation des pratiques d'évaluation sommative sera important. Ces corrélations sont détaillées et leur importance a été testée (régression) :

- *Enjeux*. Pour les épreuves de référence, la variable ENJ montre une corrélation avec tous les effets sauf HAR, mais ces liens sont moyens. Dans le contexte d'une épreuve d'orientation, les enjeux ressentis par les enseignants corréleront fortement avec les effets TTT et PRO ($r = .47$ et $.51$; $p < 1\%$).
- *Utilité*. Des corrélations modérées, voire fortes, sont aussi à noter entre l'utilité perçue et l'effet d'harmonisation des pratiques d'évaluation sommative interne et ce pour les deux types de dispositif.

- *Qualité et concordance des pratiques d'évaluation.* Plusieurs corrélations faibles sont observées entre ces perceptions de l'EE et leurs effets.

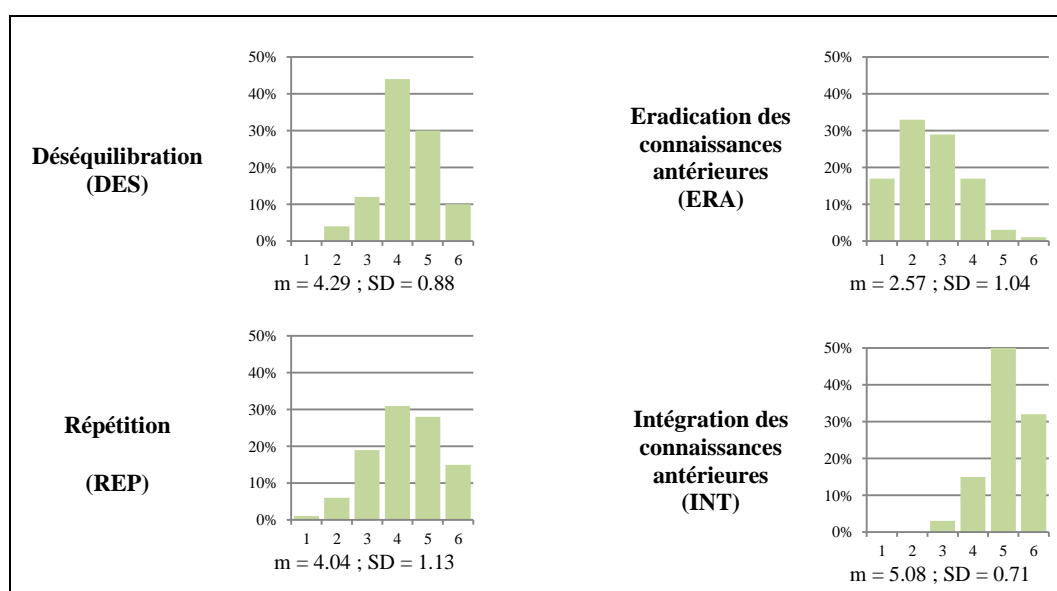
20.4. Croyances des enseignants à propos de l'enseignement

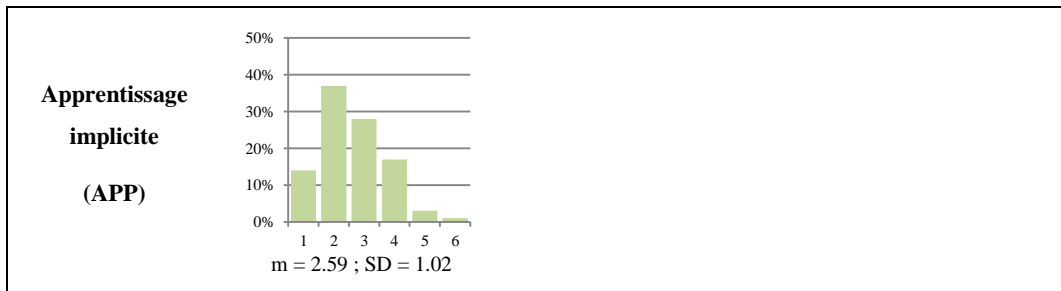
Outre leur caractéristiques individuelles et professionnelles et leur regard sur l'EE, les enseignants se différencient par des croyances diverses à propos de l'enseignement. Le questionnaire a permis d'en mesurer certaines qui, d'après la recherche exploratoire et la littérature peuvent avoir une influence sur les effets de l'EE. Il s'agit des conceptions de l'apprentissage, de l'efficacité personnelle et des conceptions de l'évaluation. Ces croyances sont tout d'abord décrites pour l'échantillon avec leurs liens. Enfin, les analyses ont permis de mesurer leur influence sur les effets de l'EE sur les pratiques des enseignants. Ces mesures ne sont pas séparées par type d'EE, les croyances des enseignants ne dépendant pas des EE auxquelles ils sont confrontés.

20.4.1. Conceptions de l'apprentissage

Le questionnaire a permis de mesurer l'adhésion ou le rejet des enseignants à certaines visions de la façon dont un élève apprend. Elles sont présentées dans le tableau ci-dessous. Les graphiques représentent la fréquence des réponses (indiquée par le nombre d'enseignants) relatives aux échelles de Likert (1 « pas du tout d'accord » à 6 « tout à fait d'accord »).

TABLEAU 27 – Description des conceptions de l'apprentissage





A la lecture du tableau 27, les enseignants interrogés acceptent en grande majorité les visions de l'apprentissage « déséquilibre », « répétitions » et « lien avec les connaissances antérieures ». Les visions « éradication des connaissances antérieures » et « apprentissage implicite » ne sont que peu reconnues. Les indices de Skewness et Kurtosis indique la bonne distribution de ces variables.

Ces variables présentent certains liens positifs ou négatifs entre elles (tableau 28).

TABLEAU 28 – Corrélations entre conceptions de l'apprentissage

	DES	REP	APP	ERA	INT
DES		.07	.04	.17**	.23**
REP			.15**	-.10	.21**
APP				.14**	-.16**
ERA					-.21**
INT					

* P<5% ; ** P<1%

Le tableau 28 montre que des corrélations significatives existent entre certaines conceptions de l'apprentissage, elles sont plutôt moyennes. Ces corrélations positives et négatives montrent des liens logiques entre visions plutôt « behavioristes » (répétition – apprentissage implicite) ou « constructivistes » (déséquilibre – intégration des connaissances antérieures), à l'exception de la corrélation entre « éradication des connaissances antérieures » et « déséquilibre » qui n'a pas de rapport théorique logique. Cette dernière reste toutefois très faible.

Une analyse en nuée dynamique (clusters) permet de créer des groupes distincts d'enseignants par rapport à leurs conceptions de l'apprentissage afin d'en dégager le profil. Dans cette analyse nous avons regroupé les enseignants en deux groupes (A et B), comme présumé par les corrélations.

- A. Acceptent les visions « constructivistes » **et ne rejettent pas** tout à fait les visions « behavioristes » de l'apprentissage (répétition, apprentissage implicite, éradication des connaissances antérieures) (N=155 ; 51%)
- B. Acceptent les visions « constructivistes » et **rejettent** les visions behavioristes de l'apprentissage (N=150 ; 49%)

Dans le tableau 29, pour chaque groupe, les scores indiquent l'éloignement des centres de clusters par rapport au centre de l'échelle de Likert (3.5). On peut ainsi observer la tendance à accepter ($m > 3.5$) ou à rejeter ($m < 3.5$) chaque dimension.

TABLEAU 29 – Clusters des conceptions de l'apprentissage

	A (N=155)	B (N=150)
Déséquilibre	+ 1	+ 0.5
Intégration des con. ant.	+ 1.5	+1.5
<i>Répétition</i>	+ 1	0
<i>Apprentissage implicite</i>	0	-1.5
<i>Eradication des con. ant.</i>	-0.5	-1.5

Les différences entre ces moyennes sont toutes significatives (sig < .01)

Pour les deux groupes, les dimensions « déséquilibre » et « intégration des connaissances antérieures » sont acceptées (comme décrit plus haut, ces deux visions font consensus pour la grande majorité des enseignants). Les deux groupes acceptent donc une vision « constructiviste » de l'apprentissage. La différence se fait sur le degré d'acceptation de la vision « behavioriste » (REP, APP, ERA).

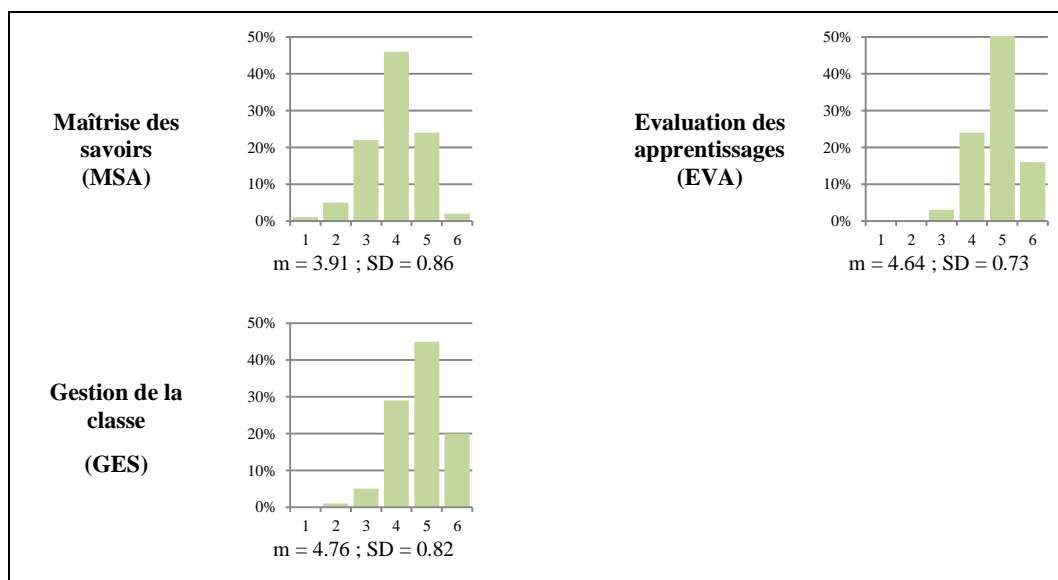
Des analyses de chi2 montrent que ces profils se différencient selon les caractéristiques suivantes :

- *Sexe*. Les hommes rejettent proportionnellement moins la vision behavioriste ($\chi^2_{(1)} = 9.95$; $p < 1\%$).
- *Formation complémentaire*. Les enseignants n'ayant pas réalisé de formation complémentaire rejettent proportionnellement moins la vision behavioriste ($\chi^2_{(1)} = 4.601$; $p < 5\%$)

20.4.2. Conceptions de l'auto-efficacité personnelle

Il a été demandé aux enseignants d'évaluer leur sentiment d'efficacité par rapport à différents domaines de l'enseignement. Les distributions des scores apparaissent dans le tableau ci-dessous. Les graphiques représentent la fréquence des réponses (indiquée par le nombre d'enseignants) relatives aux échelles de Likert (1 « pas du tout d'accord » à 6 « tout à fait d'accord »).

TABLEAU 30 – Description des conceptions de l'auto-efficacité personnelle



Le tableau 30 démontre que les enseignants qui ont répondu au questionnaire se sentent en moyenne plutôt compétents pour gérer la classe et évaluer les apprentissages des élèves. Ils se montrent plus modérés pour ce qui concerne le fait d'amener leurs élèves à maîtriser des savoirs. Il existe des liens entre ces différents domaines (tableau 31).

TABLEAU 31 – Corrélations entre domaines de l'auto-efficacité personnelle

	MSA	GES	EVA
MSA		.30**	.38**
GES			.29**
EVA			

* P<5% ; ** P<1%

Plus les enseignants se sentent compétents dans un domaine, plus ils se sentent compétents dans les autres, voici le constat à retenir du tableau 31, les corrélations étant positives et significatives (p <.01). Il est à remarquer que les domaines « maîtrise du savoir » et « évaluation » sont plus fortement corrélés. Une analyse en nuée dynamique (clusters) permet de créer des groupes distincts d'enseignant

par rapport à leurs conceptions de l'apprentissage afin d'en dégager le profil. Elle fait ressortir deux groupes (A et B).

A. Les enseignants qui se sentent **davantage compétents** dans tous les domaines (N=172 ; 56%).

B. Les enseignants qui se sentent **moins compétents** dans tous les domaines (N=133 ; 48%).

Dans le tableau 32, les scores indiquent l'éloignement des centres de clusters par rapport au centre de l'échelle de Likert (3.5). On peut donc observer pour chaque groupe la tendance à accepter ($m > 3.5$) ou à rejeter ($m < 3.5$) chaque dimension.

TABLEAU 32 – Clusters de l'efficacité personnelle

	A (N=172)	B (N=133)
MSAV	+ 1	- 0.5
GEST	+ 1.5	+ 1
EVAL	+ 1.5	+ 0.5

Les différences entre ces moyennes sont toutes significatives (sig < .01)

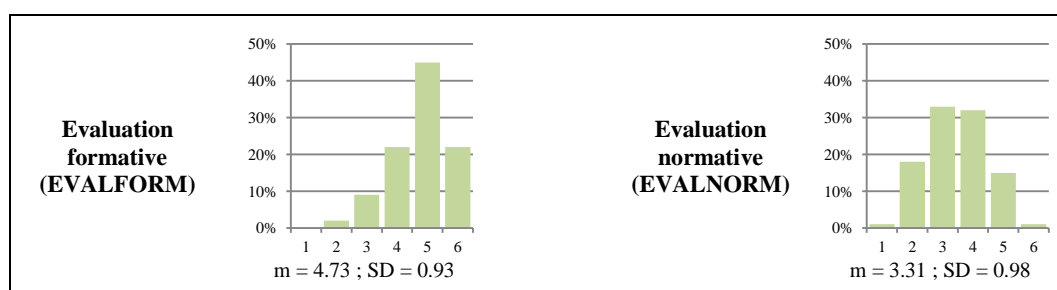
Des analyses de chi2 montrent que ces profils se différencient selon les caractéristiques suivantes :

- *Conception d'EE*. Les enseignants ayant participé à la conception d'EE se sentent proportionnellement davantage compétents ($\chi^2_{(1)} = 5.54 ; p < 5\%$).

20.4.3. Conceptions de l'évaluation scolaire

Le questionnaire a permis de mesurer l'adhésion ou le rejet des enseignants à deux visions de l'évaluation scolaire. Les graphiques représentent la fréquence des réponses (indiquée par le nombre d'enseignants) relatives aux échelles de Likert (1 « pas du tout d'accord » à 6 « tout à fait d'accord »).

TABLEAU 33 – Description des conceptions de l'évaluation scolaire



Le tableau 31 montre que les enseignants se disent dans l'ensemble davantage favorables à l'évaluation formative qu'à l'évaluation normative. Il est à noter que ces deux visions ne s'excluent pas l'une l'autre, au contraire on trouve une faible corrélation positive ($r = .24 ; p < 1\%$).

Une analyse en nuée dynamique (clusters) permet de créer des groupes distincts d’enseignant par rapport à leurs conceptions de l’apprentissage afin d’en dégager le profil. On trouve trois groupes (A, B et C)⁸⁷.

- A. Prépondérance pour l’évaluation formative **avec** rejet de l’évaluation normative (N=102 ; 33%)
- B. Prépondérance pour l’évaluation formative **sans** rejet de l’évaluation normative (N=122 ; 40%)
- C. Adhésion modérée à l’évaluation formative **avec** rejet de l’évaluation normative (N=81 ; 27%)

Dans le tableau 32, les scores indiquent l’éloignement des centres de clusters par rapport au centre de l’échelle de Likert (3.5). On peut donc observer pour chaque groupe la tendance à accepter ($m > 3.5$) ou à rejeter ($m < 3.5$) chaque dimension.

TABLEAU 34 – Clusters des conceptions de l’évaluation

	A (N=102)	B (N=122)	C (N=81)
EVALFORM	+ 2	+ 2	0
EVALNORM	- 1	+ 1	- 0.5

Il peut être conclu que la vision formative de l’évaluation est acceptée par une grande majorité des enseignants. Pour la vision normative, les enseignants sont partagés. Ces deux types d’analyses permettent en effet de constater que près de 90% des enseignants interrogés adhèrent à la vision formative de l’évaluation et que 60% d’entre eux rejettent la vision normative.

Les analyses montrent que ces profils se différencient selon les caractéristiques suivantes :

- *Sexe*. Les hommes adhèrent proportionnellement davantage à l’évaluation formative ($\chi^2(2) = 9.543$; $p < 1\%$).
- *Expérience*. Plus les enseignants sont expérimentés, moins ils rejettent la vision normative de l’évaluation ($\rho = .29$; $p < 1\%$). Notre étude étant transversale, un effet de cohorte ne peut pas être écarté.
- *Double degré*. Les enseignants qui sont en charge d’une classe à double degré adhèrent proportionnellement moins à l’évaluation normative ($\chi^2(2) = 6.05$; $p < 5\%$).

⁸⁷ On retrouve ici les mêmes groupes que Issaieva, Pini, & Crahay (2011).

Ces résultats montrent qu'il existe principalement des différences de conceptions de l'évaluation entre les générations d'enseignants du degré primaire. Les enseignants les moins expérimentés et/ou les plus fraîchement formés sont ceux qui adhèrent le plus à l'évaluation formative et rejettent le plus l'évaluation normative.

20.4.4. Liens entre croyances des enseignants à propos de l'enseignement

Comme démontré dans les résultats précédents, les différentes composantes de chaque champ de conceptions (apprentissage, efficacité, évaluation) sont liées par des corrélations positives ou négatives. Ces champs de conceptions sont donc des groupes homogènes de croyances diverses partagées ou non. Toutefois, une question souvent débattue dans la recherche est de comprendre comment fonctionnent ces croyances, comment elles se structurent (Crahay et al., 2011). Justement, les liens (corrélations) entre les composantes des divers champs de croyances peuvent être observés dans l'annexe E. Le tableau de corrélations montre que ces croyances ressemblent effectivement à des « grappes », les composantes d'un même champ de croyances corrélaient entre eux. Des corrélations significatives existent également entre des composantes de différents champs. Notamment, entre celui de l'évaluation et de l'apprentissage. Néanmoins, ces corrélations, bien que significatives restent faibles.

20.4.5. Liens entre croyances des enseignants à propos de l'enseignement et perceptions de l'EE

Les analyses démontrent qu'il existe des liens assez faibles entre croyances des enseignants à propos de l'enseignement et leurs conceptions de l'EE. En effet, les corrélations sont non significatives ou très faibles (<.25). Nous relevons tout de même des différences qui concernent les conceptions de l'évaluation scolaire :

- *EE de référence.* Les enseignants qui ne rejettent pas la vision normative de l'évaluation (groupe B) reconnaissent davantage l'utilité de l'EE pour leurs pratiques ($F_{(2,182)}= 7.612$; $p<1\%$), ce qui est confirmé par une corrélation entre les variables EVALNORM et UTI ($r=275$; $p<1\%$). Ces derniers reconnaissent également davantage la qualité des épreuves ($F_{(2,182)}=4.31$; $p<5\%$), ce qui est confirmé par la corrélation entre EVALNORM et QUA ($r=185$; $p<1\%$).

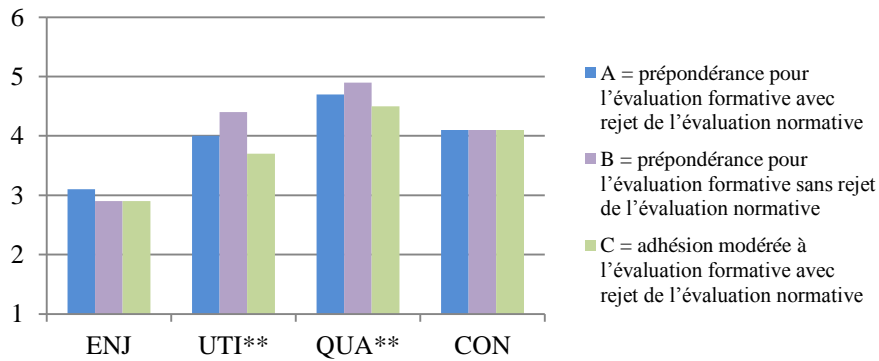


FIGURE 26 – Perceptions de l'EE et conceptions de l'évaluation scolaire (EE de référence)

- *EE d'orientation.* Les enseignants qui ne rejettent pas la vision normative de l'évaluation (groupe B) ressentent davantage d'enjeux face à l'EE ($F_{(2,110)} = 4.448$; $p < 1\%$). Ils reconnaissent également davantage la qualité de l'EE ($F_{(2,110)} = 4.965$; $p < 1\%$), ce qui est confirmé par une corrélation entre EVALNORM et QUA ($r = 0.236$; $p < 5\%$),

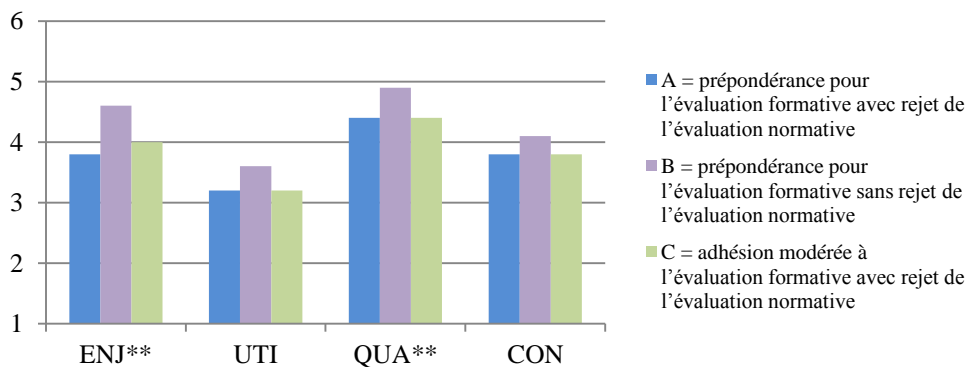


FIGURE 27 – Perceptions de l'EE et conceptions de l'évaluation scolaire (EE d'orientation)

20.4.6. Liens entre croyances des enseignants et effets de l'EE

Les analyses démontrent qu'il existe des liens entre croyances des enseignants à propos de l'enseignement (apprentissage, efficacité, évaluation) et les effets de l'EE. En effet, les corrélations sont non significatives ou très faibles ($< .25$). Nous relevons tout de même des différences, selon le type d'épreuve externe :

- *EE de référence.* Les enseignants qui appartiennent au groupe qui rejettent les visions behavioristes de l'apprentissage (groupe B) perçoivent davantage de RET ($F_{(1,183)} = 4.07$; $p < 5\%$), davantage de PRO ($F_{(1,183)} = 7.71$; $p < 1\%$), de COL ($F_{(1,183)} = 4.539$; $p < 5\%$) et de HAR ($F(1) = 3.96$; $p < 5\%$).

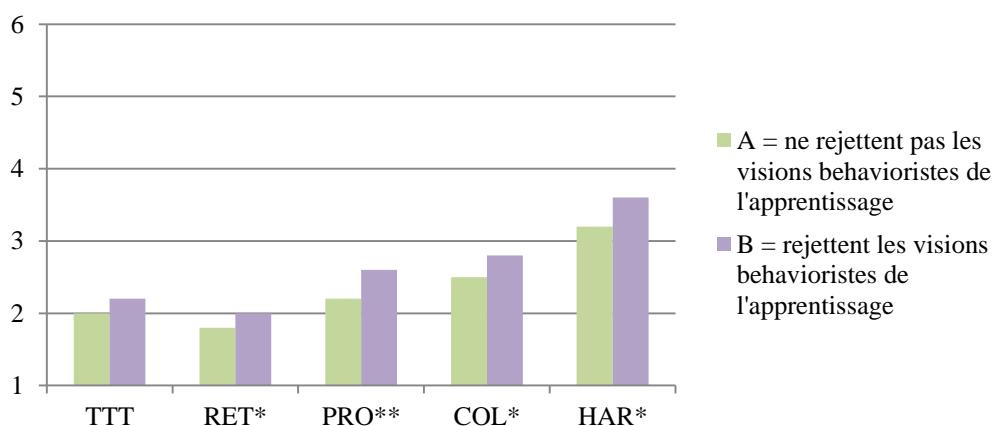


FIGURE 28 - Effets de l'EE et conceptions de l'apprentissage

Egalement, Les enseignants qui se sentent le moins efficaces affirment ressentir davantage de RET ($F_{(1,181)} = 10.92$; $p < 1\%$).

- *EE d'orientation.* Les enseignants qui rejettent la vision normative de l'évaluation (groupe B) perçoivent davantage de TTT ($F_{(2,181)} = 4.47$; $p < 5\%$). Il existe une corrélation positive entre TTT et EVALNORM ($r = .31$; $p < 1\%$).

20.5. Modélisation des variables qui influencent les perceptions de l'EE

Des analyses de régression linéaires sont réalisées dans le but d'expliquer la variance des perceptions de l'EE. Pour chaque dimension, un même modèle est testé. Il comprend les variables suivantes, regroupées en 5 dimensions :

- 1) Le profil professionnel de l'enseignant : EXP ;
- 2) Les perceptions de l'EE de l'enseignant : ENJ, UTI, QUA, CON ;
- 3) Les conceptions de l'enseignant à propos de l'apprentissage : DES, REP, APP, ERA, INT ;
- 4) Les conceptions de l'enseignant à propos de son auto-efficacité personnelle : MSA, GES, EVA ;
- 5) Les conceptions de l'enseignant à propos de l'évaluation scolaire : EVALFORM, EVALNORM.

Les analyses de régression permettent de mesurer l'influence de ces variables sur les perceptions de l'EE. Pour ces analyses de régression, la méthode *enter* a été choisie, elle permet d'inclure tous les prédicteurs dans le modèle. Pour chaque analyse, le Variance Inflation Factor (VIF) a permis de s'assurer que la corrélation entre les variables n'est pas trop forte ($VIF < 10$). Dans le tableau 35, les

coefficients standardisés (β) donnent le poids relatif de la variable dans le modèle et le sens de la relation (positive ou négative). La signification (p) indique si la variable contribue significativement ou non à expliquer la variance de la variable dépendante. Enfin, la variance totale expliquée par le modèle est indiquée par le coefficient R^2 . Aussi, afin d'observer le pourcentage de variance expliquée par chacune des 5 dimensions (R^2), des analyses de régression ont été testées séparément⁸⁸. A la suite de ce tableau, les liens significatifs sont schématisés et commentés pour chaque perception de l'EE.

TABLEAU 35 – Analyses de régression pour les perceptions de l'EE

		ENJ		UTI		QUA		CON	
		référence	orientation	référence	orientation	référence	orientation	référence	orientation
1)	EXP	-.07	-.17	.04	.07	.10	.06	.10	-.04
	R^2	0%	0%	0%	0%	0%	0%	0%	0%
2)	ENJ			.18**	-.08	-.09	-.11	-.05	-.03
	UTI	.15	-.14			.40**	.17	.19**	.30**
	QUA	-.05	.04	.32**	.15			.12	.12
	CON	-.05	-.19**	.14	.31**	.14	.17		
	R^2	3%	4%	21%	11%	20%	6%	7%	11%
	DES	-.01	-.14	-.04	-.06	.17*	-.06	.07	.15
3)	REP	.14	.05	.04	-.01	.07	.05	-.01	-.07
	APP	-.07	-.05	-.07	-.08	.06	.20*	.11	.15
	ERA	-.05	.03	.14*	.04	.00	.14	.04	.17
	INT	.00	-.02	.07	.04	.05	-.04	.02	-.11
	R^2	1%	1%	4%	1%	3%	6%	1%	7%
	MSA	-.09	-.12	.18**	-.04	.12	.08	.21**	-.11
4)	GES	-.21*	-.04	-.13	-.13	.12	.07	-.07	-.08
	EVA	-.15	.14	-.05	-.02	.01	.06	.08	.02
	R^2	4%	2%	2%	3%	1%	1%	6%	0%
	EVALFORM	.07	-.01	.14	.13	.03	.18	.04	.01
5)	EVALNORM	-.15	.13	.21**	.15	.06	.28**	-.05	.02
	R^2	0%	0%	5%	0%	0%	5%	0%	0%
R^2 total		8%	8%	27%	15%	24%	17%	14%	18%

**p<1% *p<5%

⁸⁸ La variance totale expliquée n'est pas toujours égale à l'addition de celle expliquée par les dimensions, la multicollinéarité n'est jamais nulle et peut modifier cette valeur.

20.5.1. Enjeux ressentis face à l'EE

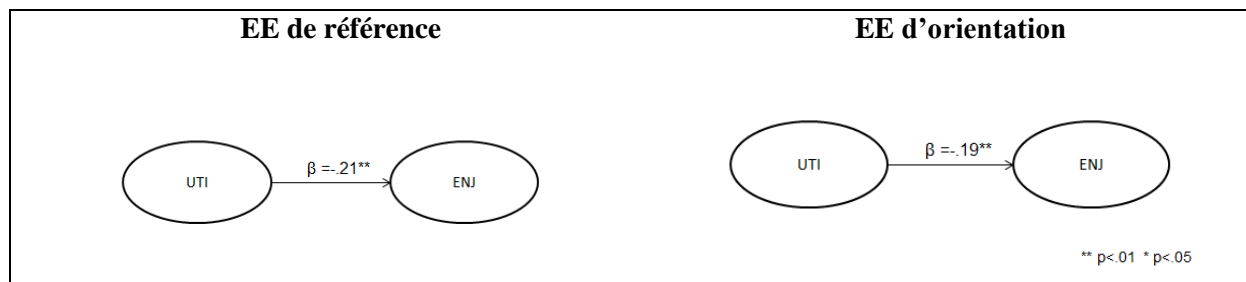


FIGURE 29 – Les variables qui ont une influence sur ENJ

EE de référence. Les analyses de régression montrent que, des variables testées, seule la gestion de la discipline a une influence sur les enjeux ressentis ($\beta = -.21$). La relation est négative, signifiant que moins les enseignants se sentent compétents dans ce domaine, plus ils ressentent d'enjeux. Le modèle testé est significatif ($F_{(1,175)}=7.82$; $p < 1\%$), les prédicteurs expliquent 8% de la variance observée pour cet effet ($R^2 = .08$).

EE d'orientation. Pour ce type d'EE, seule la concordance des pratiques d'évaluation interne et externe a une influence sur les enjeux ressentis ($\beta = -.19$). La relation est négative, signifiant que plus les pratiques concordent, moins ils ressentent d'enjeux. Le modèle testé est significatif ($F_{(1,105)}=4.06$; $p < 1\%$), les prédicteurs expliquent 4% de la variance observée pour cet effet ($R^2 = .04$).

20.5.2. Utilité de l'EE

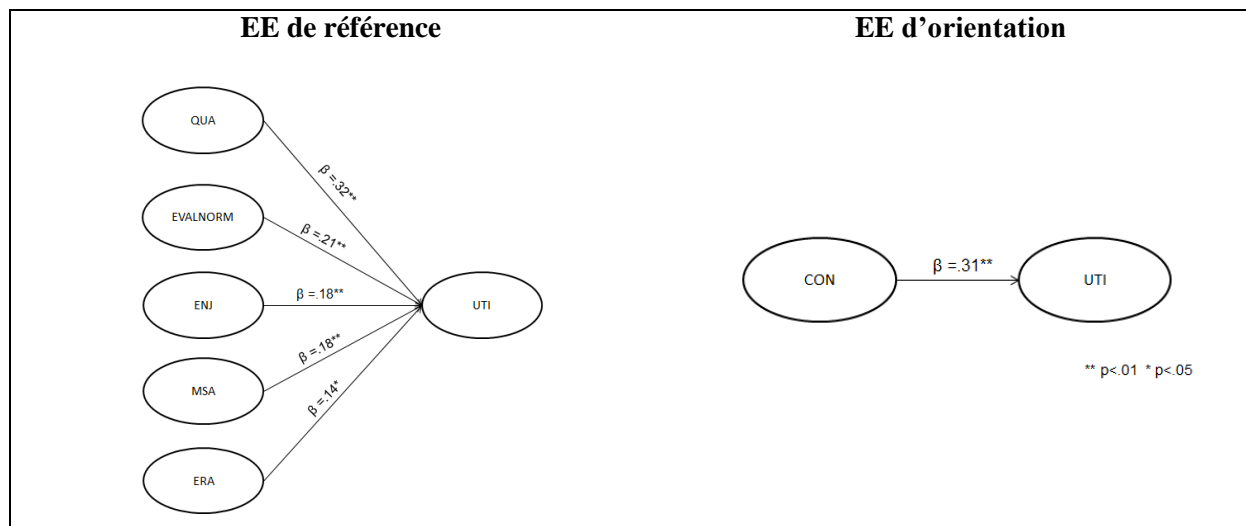


FIGURE 30 – Les variables qui ont une influence sur UTI

EE de référence. Les analyses de régression montrent que quatre variables testées ont une influence sur l'utilité de l'EE perçue par l'enseignant, dans l'ordre d'importance: la qualité de l'EE ($\beta = .32$), la vision normative de l'évaluation scolaire ($\beta = .21$), les enjeux ($\beta = .18$), la maîtrise des savoirs ($\beta = .18$)

et la vision de l'apprentissage comme une éradication des connaissances antérieures ($\beta = .14$). Le modèle testé est significatif ($F_{(5,171)}=14.42$; $p<1\%$), les prédicteurs expliquent 8% de la variance observée pour cet effet ($R^2 = .08$).

EE d'orientation. Pour ce type d'EE, seul la concordance des pratiques d'évaluation interne et externe a une influence sur l'utilité de l'EE ($\beta = .31$). La relation est positive, signifiant que plus les pratiques évaluatives internes et externes concordent, plus les enseignants voient l'utilité de l'EE pour leurs pratiques. Le modèle testé est significatif ($F_{(1,105)}=11.08$; $p<1\%$), les prédicteurs expliquent 27% de la variance observée pour cet effet ($R^2 = .27$).

20.5.3. Qualité de l'EE

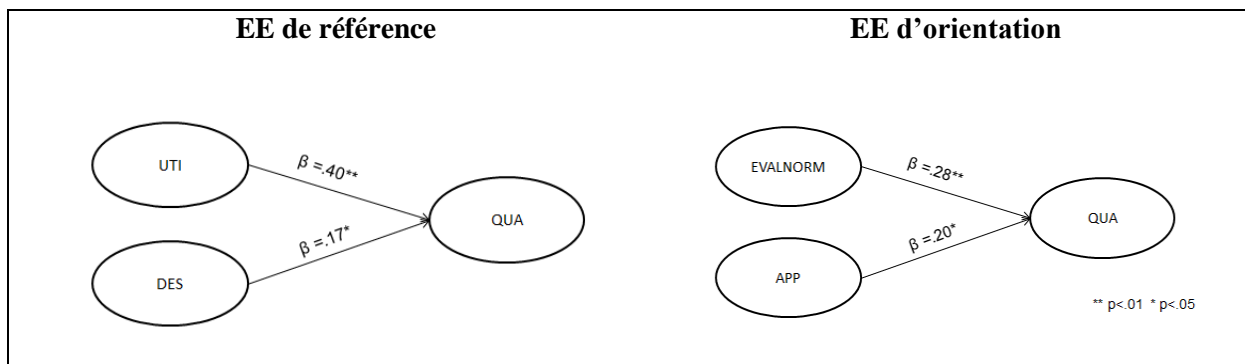


FIGURE 31 – Les variables qui ont une influence sur QUA

EE de référence. Les analyses de régression montrent que, deux variables testées ont une influence sur la qualité de l'EE perçue par l'enseignant, dans l'ordre d'importance: l'utilité de l'EE ($\beta = .40$), la vision de l'apprentissage comme une déséquilibration ($\beta = .17$). Le modèle testé est significatif ($F_{(5,171)}=14.42$; $p<1\%$), les prédicteurs expliquent 24% de la variance observée pour cet effet ($R^2 = .24$).

EE d'orientation. Les analyses de régression montrent que, deux variables testées ont une influence sur la qualité de l'EE perçue par l'enseignant, dans l'ordre d'importance: la vision normative de l'évaluation scolaire ($\beta = .28$), la vision de l'apprentissage comme implicite ($\beta = .20$). Le modèle testé est significatif ($F_{(1,105)}=11.08$; $p<1\%$), les prédicteurs expliquent 17% de la variance observée pour cet effet ($R^2 = .17$).

20.5.4. Concordance des pratiques d'évaluation interne et externe

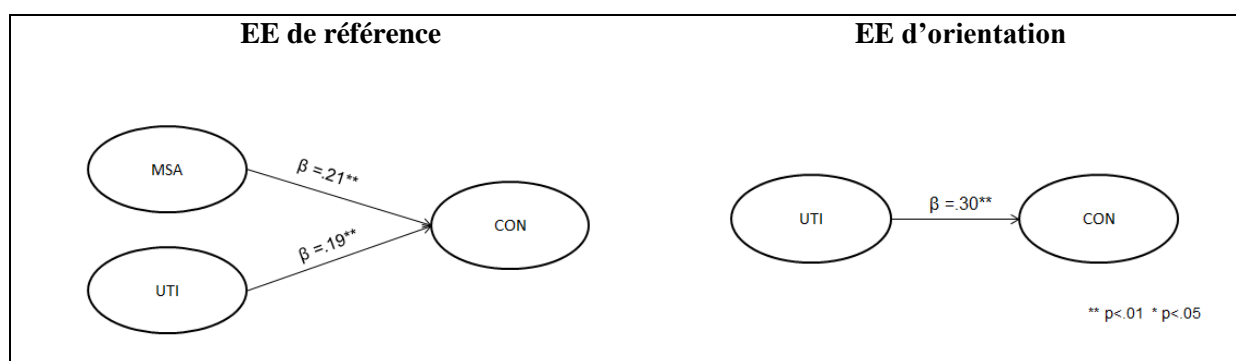


FIGURE 32 – Les variables qui ont une influence sur CON

EE de référence. Les analyses de régression montrent que, deux prédicteurs testés ont une influence sur cette variable, dans l'ordre d'importance: la maîtrise des savoirs ($\beta = .21$), l'utilité de l'EE ($\beta = .19$). Le modèle testé est significatif ($F_{(2,175)}=14.42$; $p < 1\%$), les prédicteurs expliquent 14% de la variance observée pour cet effet ($R^2 = .14$).

EE d'orientation. Pour ce type d'EE, seule l'utilité de l'EE a une influence sur cette variable ($\beta = .30$). Le modèle testé est significatif ($F_{(1,109)}=10.98$; $p < 1\%$), les prédicteurs expliquent 18% de la variance observée pour cet effet ($R^2 = .28$).

20.6. Modélisation des variables qui influencent les effets de l'EE sur les pratiques des enseignants

Des analyses de régression linéaires sont réalisées dans le but d'expliquer la variance des effets de l'EE. Pour chaque effet, un même modèle est testé. Il comprend les variables suivantes, regroupées en 5 dimensions :

- 1) *Le profil professionnel de l'enseignant* : EXP ;
- 2) *Les perceptions de l'EE de l'enseignant* : ENJ, UTI, QUA, CON ;
- 3) *Les conceptions de l'enseignant à propos de l'apprentissage* : DES, REP, APP, ERA, INT ;
- 4) *Les conceptions de l'enseignant à propos de son auto-efficacité personnelle* : MSA, GES, EVA ;
- 5) *Les conceptions de l'enseignant à propos de l'évaluation scolaire* : EVALFORM, EVALNORM.

Les analyses de régression permettent de mesurer l'influence de ces variables sur les effets de l'EE. La procédure mise en œuvre est la même que dans le point précédent.

TABLEAU 36 – Analyses de régression pour les effets de l’EE

		TTT		RET		PRO		COLL		HAR	
		référence	orientation	référence	orientation	référence	orientation	référence	orientation	référence	orientation
1)	EXP	.02	.05	-.02	-.01	.23**	-.02	-.02	-.14	.04	.06
	R ²	0%	0%	0%	0%	5%	1%	0%	0%	0%	0%
2)	ENJ	.26**	.46**	.23**	.50**	.27**	.23*	.20**	.08	.05	.08
	UTI	.08	.02	.06	-.01	.15*	.26**	.04	.20	.45**	.46**
	QUA	-.04	.18*	-.05	.04	.03	-.05	-.02	.01	.05	.08
	CON	-.06	-.02	-.04	-.08	-.10	-.19*	-.04	-.17	.08	-.05
	R ²	9%	30%	7%	26%	11%	9%	5%	5%	22%	24%
3)	DES	.11	.09	.07	-.03	.07	-.15	-.00	.09	.06	.10
	REP	.18*	.09	.01	.04	.06	.10	.10	.04	-.18	-.07
	APP	.14*	-.02	.21**	.07	.23**	.09	.13	.16	.11	-.05
	ERA	.03	.01	-.04	-.08	.11	.21*	.10	.13	.08	.18*
	INT	-.01	.09	-.08	-.04	.02	.06	-.08	.06	.10	.06
	R ²	7%	5%	5%	2%	8%	8%	7%	6%	7%	7%
4)	MSA	-.05	.02	-.18*	-.13	-.09	.04	.03	-.06	.12	.03
	GES	-.02	-.12	-.08	-.15	-.05	-.01	.07	.05	.04	.01
	EVA	-.01	.08	-.08	-.14	-.14	-.11	.08	-.01	.02	-.07
	R ²	1%	3%	6%	4%	2%	2%	1%	0%	1%	1%
5)	EVALFORM	-.00	-.03	-.04	.01	.07	-.05	.10	.03	.11	-.15
	EVALNORM	.08	.24**	.09	.10	.09	.06	.20**	.06	-.10	-.08
	R ²	1%	10%	0%	4%	2%	0%	6%	1%	3%	1%
R² total		18%	37%	18%	31%	27%	20%	16%	12%	27%	29%

**p<1% *p<5%

20.6.1. Teaching to the test

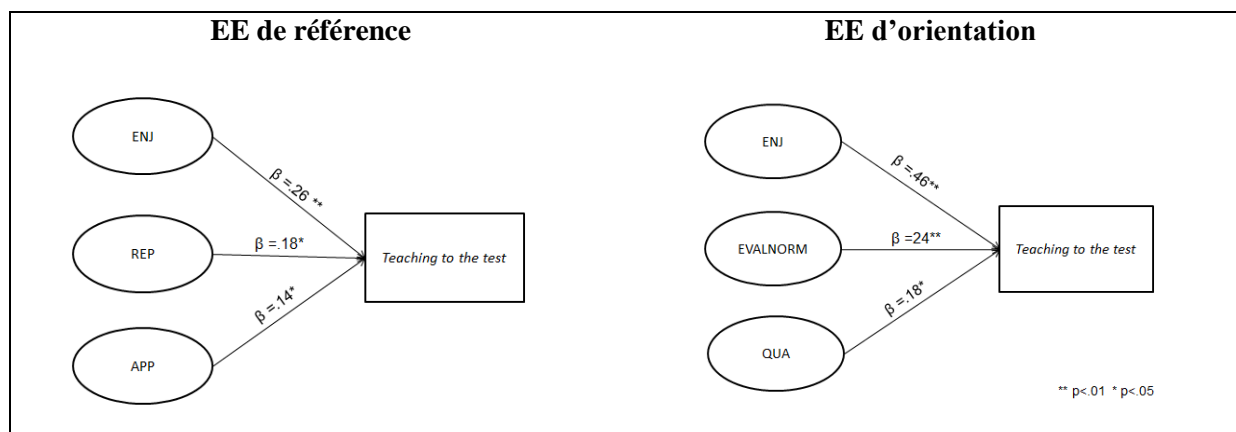


FIGURE 33 – Les variables qui ont une influence sur l’effet TTT

EE de référence. Les analyses de régression montrent que, des variables testées, les enjeux ressentis par les enseignants face à l’EE ont l’influence la plus importante sur l’effet de TTT ($\beta = .26$). La relation est positive, signifiant que plus les enseignants ressentent d’enjeux, plus ils rapportent de TTT. Les conceptions de l’apprentissage des enseignants ont également une influence sur cet effet, notamment les variables qui se rapportent à des visions plutôt behavioristes de l’apprentissage,

« répétition » ($\beta = .18$) et « apprentissage implicite » ($\beta = .14$). Le modèle testé est significatif ($F_{(3,176)}=8.95$; $p<1\%$), les prédicteurs expliquent 18% de la variance observée pour cet effet ($R^2 = .18$).

EE d'orientation. Pour ce type d'EE, les enjeux ressentis par les enseignants (ENJ) occupent un rôle prépondérant dans la prédiction de cet effet ($\beta = .46$). Les conceptions d'évaluation normative (EVALNORM) ($\beta = .24$) et la qualité perçue de l'EE ($\beta = .18$) ont également une valeur prédictive forte. Le modèle est significatif ($F_{(3,103)}=12.17$; $p<1\%$), les prédicteurs expliquent 37% de la variance observée ($R^2 = .37$).

20.6.2. Rétrécissement du curriculum

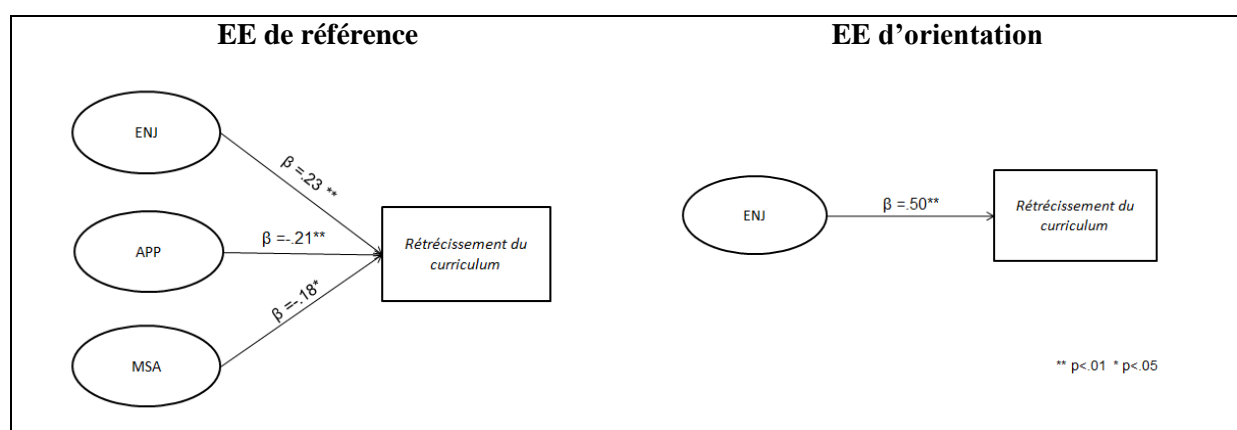


FIGURE 34 – Les variables qui ont une influence sur l'effet RET

EE de référence. Les enjeux face à l'EE ont une influence importante ($\beta = .23$) sur cet effet et les variables APP ($\beta = .21$) et MSAV ($\beta = -.18$) moindre. Il est à remarquer l'effet négatif de cette dernière variable, ce qui signifie que moins les enseignants se sentent compétents pour amener leurs élèves à maîtriser les savoirs, plus ils vont rapporter l'effet de RET. Le modèle présenté est significatif ($F_{(3,179)}=9.08$; $p<1\%$), les prédicteurs expliquent 18% de la variance observée ($R^2 = .18$).

EE d'orientation. La variable ENJ prend un rôle prépondérant ($\beta = .50$) dans l'explication de la variance de cet effet. La variable MSA n'entre pas dans ce modèle, malgré les liens démontrés auparavant (ils étaient toutefois faibles, ceci n'est pas dû à un effet médiateur, les deux variables n'ayant aucun lien). Ce résultat montre que les enjeux ressentis par les enseignants sont la principale cause du rétrécissement du curriculum enseigné pour une épreuve d'orientation. Le modèle présenté est significatif ($F_{(1,106)}=34.56$; $p<1\%$), les prédicteurs expliquent 31% de la variance observée ($R^2 = .31$).

20.6.3. Développement du professionnalisme

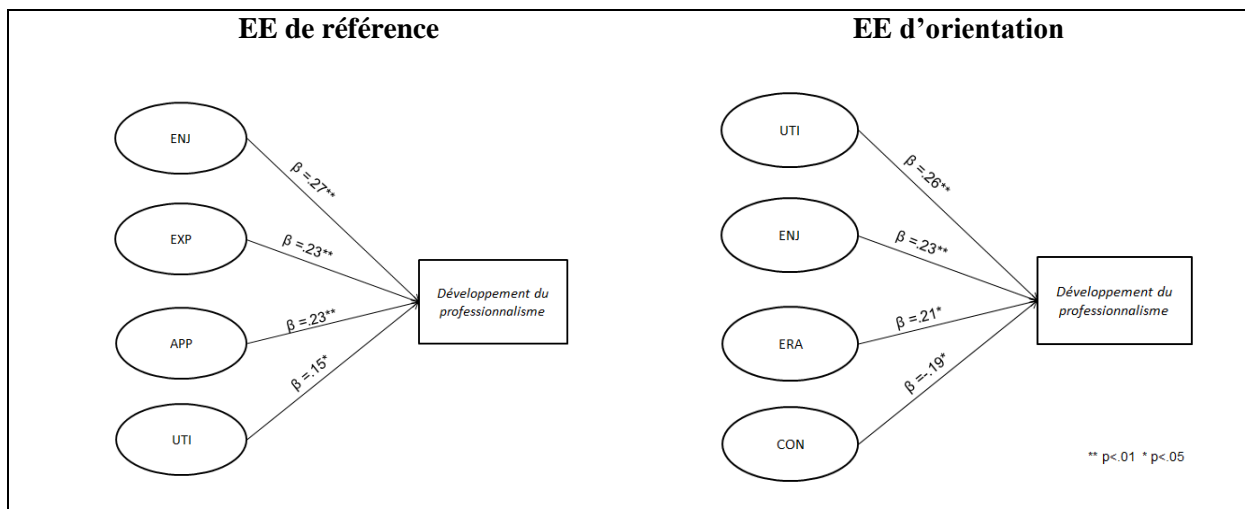


FIGURE 35 – Les variables qui ont une influence sur l’effet PRO

EE de référence. Cet effet est expliqué par que variables : ENJ ($\beta = .27$), EXP ($\beta = .23$), APP ($\beta = .23$) et « UTI » ($\beta = .15$). Le modèle est présenté significatif ($F_{(4,172)}=11.78$; $p<1\%$) et explique 27% de la variance observée ($R^2 = .27$).

EE d’orientation. Cet effet est expliqué par quatre variables : UTI ($\beta = .26$), ENJ ($\beta = .23$), ERA ($\beta = .21$) et CON ($\beta = -.19$). Le modèle présenté est significatif ($F_{(4,102)}=4.55$; $p<1\%$) et explique 20% de la variance observée ($R^2 = .20$).

20.6.4. Collaboration

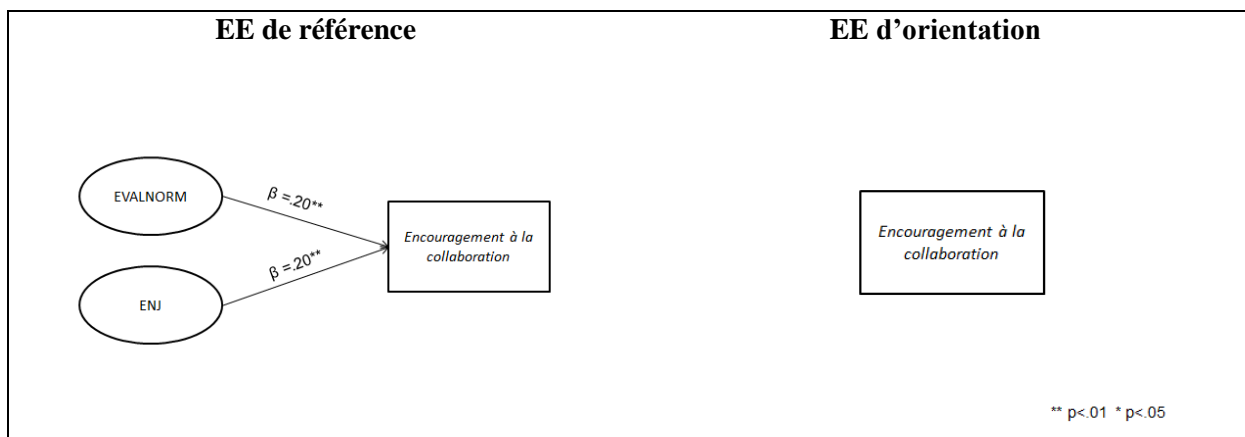


FIGURE 36 – Les variables qui ont une influence sur l’effet COL

EE de référence. Cet effet est expliqué par deux variables : EVALNORM ($\beta = .20$) et ENJ ($\beta = .20$). Ce modèle est significatif ($F_{(2,174)}=9.02$; $p<1\%$) et explique 16% de la variance observée ($R^2 = .16$).

EE d’orientation. Aucune variable n’explique la variance de cet effet pour ce type d’EE.

20.6.5. Harmonisation des pratiques d'évaluation sommative

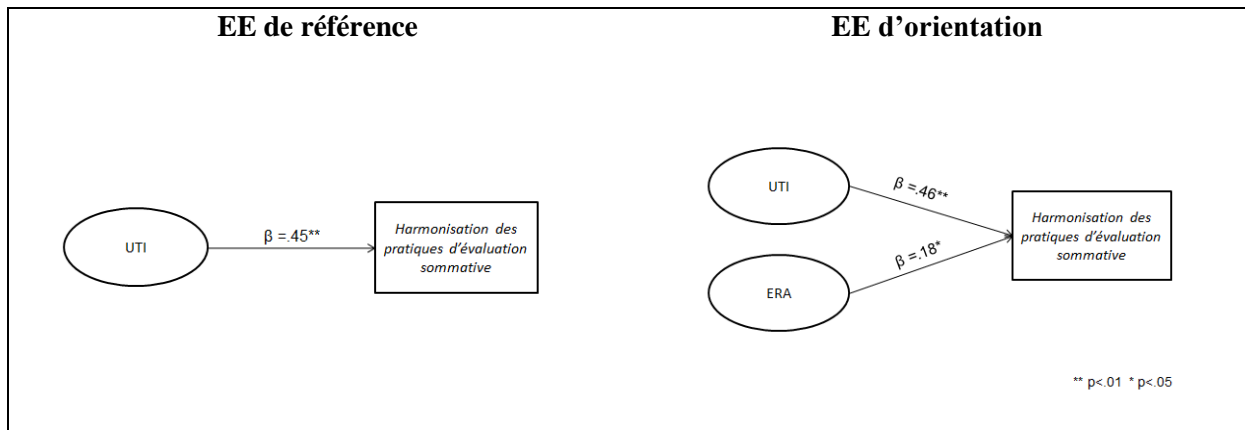


FIGURE 37 – Les variables qui ont une influence sur l'effet HAR

EE de référence. Cet effet est expliqué par une seule variable, UTI ($\beta = .45$). Ce modèle est significatif ($F_{(1,176)}=45.75$; $p < 1\%$) explique 27% de la variance observée ($R^2 = .27$).

EE d'orientation. Cet effet est expliqué par deux variables : UTI ($\beta = .46$) et ERA ($\beta = .18$). Ce modèle est significatif ($F_{(1,176)}=45.75$; $p < 1\%$) explique 29% de la variance observée ($R^2 = .29$).

CHAPITRE 5

INTERPRÉTATION DES RÉSULTATS

21. Les perceptions de l'évaluation externe par les enseignants (QR1)	204
22. Les effets de l'évaluation externe sur les pratiques des enseignants (QR2)	209
23. Les facteurs qui influencent les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants (QR3).....	213

21. Perceptions de l'évaluation externe par les enseignants (QR1)

Nous reprenons et interprétons dans ce point les résultats des deux parties empiriques qui ont trait à la première question de recherche (QR1). Nous vérifions tout d'abord l'acceptation ou le rejet des hypothèses qui lui sont sous-jacentes, puis nous présentons une synthèse sous forme de réponse à la question de recherche. Ces résultats sont discutés dans le chapitre 6 en les confrontant à la littérature.

QR1

Comment les enseignants du primaire perçoivent-ils l'évaluation externe des acquis des élèves ?

H1.1

Les enseignants sont favorables à l'évaluation externe des acquis des élèves mais craignent l'usage de ses résultats.

Les enseignants interrogés dans la première partie empirique de la thèse ont un regard contrasté quant à l'évaluation externe. Seul un enseignant dit être totalement opposé à ce genre de dispositif (Claude ; p.123). Dans l'analyse des entretiens deux dimensions ont été principalement relevées quant aux perceptions de l'évaluation externe (p.142), à des degrés différents pour chaque enseignant :

- *Estime du modèle de l'EE.* Pour les enseignants interrogés (à une exception près), les épreuves externes servent avant tout à valider leur travail. Ces dispositifs leur donnent des informations "fiables" pour vérifier la pertinence de leurs propres pratiques d'enseignement et d'évaluation. L'évaluation externe leur donne aussi des indications sur les contenus du programme. Par contre, pour un seul enseignant (Claude), l'évaluation est totalement inutile pour son métier, ses propres pratiques lui suffisent. Tous les enseignants interrogés relèvent néanmoins la bonne qualité des épreuves.
- *Enjeux liés aux résultats des élèves.* Tous les enseignants interrogés relèvent le même inconvénient des épreuves externes. Il est lié aux résultats des élèves (s'agissant ici d'une épreuve d'orientation). En effet, les enjeux liés aux résultats sont trop importants pour les élèves et leurs parents (orientation) et pour les enseignants (validation de leur pratique ; contrôle "indirect" de leur travail). Les enseignants ressentent donc un certain stress. De plus, les délais très courts (lorsque l'épreuve a lieu en cours d'année scolaire et non à la fin) mettent une pression supplémentaire sur les épaules de enseignants qui doivent parvenir à traiter tous les contenus à temps.

La seconde phase empirique a permis de mesurer les perceptions des enseignants quant à l'EE de manière plus détaillée, et ce pour les deux types d'évaluation externe investigués (de référence et d'orientation)⁸⁹. En effet, dans cette deuxième phase empirique, la dimension « estime du modèle de l'EE » a été divisée en deux dimensions « utilité perçue de l'EE » et « qualité de l'EE ». La dimension « concordance des pratiques d'évaluation » a également été ajoutée (tableau 24 ; p.181) :

- *Utilité perçue de l'EE.* Les enseignants interrogés perçoivent les deux types d'EE comme plutôt utiles pour leurs pratiques.
- *Qualité perçue de l'EE.* Les enseignants interrogés reconnaissent en majorité la qualité de chaque type d'EE.
- *Enjeux ressentis face à l'EE.* Les enseignants interrogés ressentent des enjeux face aux deux types d'EE.
- *Concordance des pratiques d'évaluation.* Les enseignants interrogés reconnaissent que les pratiques évaluatives des épreuves externes concordent avec les leurs.

Les analyses statistiques montrent également des liens entre certaines perceptions (tableau 25 ; p.183). Pour chaque type d'EE, plus les enseignants observent une concordance entre leurs pratiques évaluatives et celles de l'EE, plus ils reconnaissent l'utilité de l'EE. La relation est également positive entre leurs perceptions de la qualité et de l'utilité de l'EE.

L'hypothèse est confirmée, les enseignants reconnaissent l'utilité et la qualité de l'évaluation externe des acquis des élèves, mais craignent l'usage de leurs résultats (ils redoutent leurs enjeux). Toutefois, ces résultats sont variables selon le type d'EE et entre les enseignants.

H1.2

Les perceptions des enseignants quant à l'évaluation externe des acquis des élèves dépendent des finalités de l'épreuve externe.

Note étude permet de comparer les perceptions des enseignants quant à deux types d'épreuve externe. Ces deux types d'EE se différencient par la finalité et à l'usage de leurs résultats (de référence ou d'orientation). Seule la partie quantitative de ce travail permet de vérifier cette hypothèse. Premièrement, la partie exploratoire n'implique que des enseignants confrontés à l'EE d'orientation. Deuxièmement, seules les analyses statistiques permettent de quantifier les perceptions et donc de les comparer pour les deux types d'EE.

⁸⁹ Les différences sont traitées dans l'hypothèse H1.2 (p.205-206).

Premièrement, les résultats issus du questionnaire (tableau 24 ; p.181) nous renseignent sur la nature des perceptions que les enseignants ont de ces deux types d'épreuve externe⁹⁰.

- *EE de référence.* L'évaluation externe de référence entraîne en moyenne des enjeux faibles à modérés pour les enseignants (m=3.01). Ceci peut s'expliquer par le fait que les résultats de ce type d'EE n'ont pas, ou peu, d'impact sur la carrière scolaire des élèves. Toutefois, ils valident tout de même les pratiques des enseignants de manière interne (l'enseignant peut comparer les résultats de sa classe à une norme). Ce type d'épreuve est reconnu comme utile par les enseignants pour leurs pratiques (m=4.14). Selon les items, elle donne aux enseignants la possibilité de vérifier et de modifier leurs pratiques. La qualité de ce type d'EE est aussi reconnue (m=4.71). Enfin, pour les enseignants, leur pratiques évaluatives concordent plutôt avec celles de l'EE (m=4.08).
- *EE d'orientation.* L'EE d'orientation entraîne en moyenne des enjeux modérés à forts pour les enseignants (m=4.31). La pression ressentie par les enseignants vient du fait que les résultats ont un impact sur la carrière scolaire des élèves (orientation) et valident les pratiques des enseignants de manière interne et externe (les inspecteurs et les directeurs de l'école secondaire ont accès aux résultats des élèves ; une comparaison est faite entre les résultats des épreuves et l'évaluation sommative des enseignants). Ce type d'épreuve est reconnu en moyenne comme modérément utile par les enseignants pour leurs pratiques (m=3.40). La qualité de ces épreuves est aussi reconnue (m=4.61). Enfin, pour les enseignants, leurs pratiques évaluatives concordent plutôt avec celles de l'EE (m=3.96).

Les analyses statistiques permettent d'observer les différences entre les perceptions de l'EE des enseignants entre les deux types d'EE (figure 25 ; p.182). Plusieurs constats se dégagent de ces résultats. L'EE d'orientation entraîne significativement davantage d'enjeux. L'EE de référence quant à elle est perçue comme davantage utile par les enseignants. Enfin, les dimensions de qualité et de concordance des pratiques d'évaluation interne et externe ne diffèrent pas significativement entre les deux types d'EE.

L'hypothèse est donc partiellement confirmée. Si les caractéristiques "techniques" de l'EE (qualité, concordance) font consensus entre les deux types d'EE, leurs caractéristiques liées à l'usage (interne et externe à la classe) de leurs résultats créent des différences. L'EE d'orientation est perçue par les enseignants comme entraînant des enjeux plus forts que l'EE de référence. L'EE de référence est perçue comme davantage utile que l'EE d'orientation.

⁹⁰ Les échelles de mesure de type « Likert » montrent le degré d'accord des répondant de 1 (pas du tout d'accord) à 6 (tout à fait d'accord). Nous considérons les mesures, selon leur moyenne, de la manière suivante: < 3 = faible ; 3 à 4 = modéré ; > 4 = fort. Les items ont été construits dans cette optique.

H1.3

Les perceptions des enseignants quant à l'évaluation externe des acquis des élèves dépendent de leur expérience professionnelle.

L'analyse des entretiens montre certaines tendances, bien que portant sur des situations singulières. En effet, certains récits montrent que les enseignants perçoivent l'EE différemment selon qu'ils ont plus ou moins d'expérience professionnelle. Ainsi, Vincent (p.125) qui est dans ses premières années d'enseignement affirme que son manque d'expérience amène une certaine pression quant aux résultats de ses élèves à l'épreuve externe. Il doit prouver qu'il est compétent. Il donne également une grande utilité à l'EE, dans un but d'amélioration de ses pratiques. Avec un degré d'expérience analogue à celui de Vincent, Patrick (p.132) n'a pas les mêmes perceptions que son collègue ; il travaille toutefois en étroite collaboration avec des enseignants plus expérimentés. A l'inverse, certains enseignants plus expérimentés ne trouvent que peu (Ghislain ; p.140) ou pas du tout (Claude ; p.122) d'utilité à l'EE.

L'éventuelle influence de l'expérience professionnelle sur les perceptions de l'EE a pu être mesurée plus clairement dans les analyses quantitatives. Celles-ci montrent que les jeunes enseignants (moins de 10 années d'expérience professionnelle) ressentent davantage d'enjeux pour l'EE d'orientation (p.183). Toutefois, aucune relation ni différence significative n'est observée entre l'expérience et les perceptions de l'EE. L'expérience ne ressort pas non plus comme un prédicteur significatif des perceptions dans les analyses de régression (tableau 35 ; p.194).

L'hypothèse n'est donc pas confirmée. Les perceptions des enseignants quant à l'évaluation externe des acquis des élèves ne dépendent pas directement de leur expérience professionnelle.

H1.4

Les perceptions des enseignants quant à l'évaluation externe des acquis des élèves dépendent de leurs croyances à propos de l'enseignement (conceptions de l'apprentissage, de l'évaluation scolaire et de leur auto-efficacité personnelle).

Même si, lors des entretiens, les enseignants ont été interrogés à propos de leurs conceptions de l'évaluation scolaire (en général), les analyses qualitatives ne permettent pas de vérifier cette hypothèse.

Par contre, quelques liens (corrélations), plutôt disparates sont observés dans la partie quantitative (p.191-193). Les conceptions de l'évaluation scolaire ont une certaine influence sur les perceptions de l'EE. Les résultats montrent que les enseignants qui ne rejettent pas la vision normative de l'évaluation scolaire voient davantage d'utilité aux épreuves externes (la différence n'est pas significative pour l'EE d'orientation) et leur trouve davantage de qualités. Ces liens sont confirmés par les analyses de

régression (tableau 35 ; p.194). Toutefois, le poids de ce prédicteur reste faible. Ce groupe d'enseignants ressent également davantage d'enjeux aux EE d'orientation.

Cette relation entre vision normative de l'évaluation scolaire et perceptions de l'EE peut s'expliquer de deux manières. Premièrement, les enseignants qui voient l'évaluation scolaire comme un moyen de comparer les élèves entre eux peuvent voir une plus grande utilité et une plus grande qualité à ces épreuves qui sont de fait normative (les deux types d'EE permettent de placer les résultats des élèves en rapport à une norme cantonale). Deuxièmement, elles permettent à ces enseignants de classer leur propre performance en comparant les résultats de leur classe à ceux des autres classes. Ce dernier élément explique le fait qu'ils ressentent également davantage d'enjeux.

L'hypothèse n'est pas confirmée, mis à part de faibles liens avec les conceptions de l'évaluation, les perceptions des enseignants quant à l'évaluation externe des acquis des élèves ne dépendent que faiblement des croyances des enseignants à propos de l'enseignement.

Synthèse QR1 : Comment les enseignants du degré primaire perçoivent-ils l'évaluation externe des acquis des élèves ?

Les résultats aussi bien qualitatifs que quantitatifs montrent que les enseignants ont des perceptions diverses de l'EE, parfois contradictoires. Ils montrent qu'elles sont perçues par ceux-ci comme un "mal nécessaire", étant utiles pour leurs pratiques (valider et modifier leurs pratiques) mais en même temps elles leur font ressentir une certaine pression liée aux résultats.

Sous l'angle de leurs caractéristiques "techniques" (qualité, concordance), les deux types d'EE ne se distinguent pas. En revanche, sous celui de l'usage de leurs résultats (interne et externe à la classe) se créent des différences. L'EE d'orientation est perçue par les enseignants comme entraînant davantage d'enjeux que l'EE de référence. A l'inverse, l'EE de référence est perçue comme davantage utile que l'EE d'orientation.

Les analyses ont pu démontrer que ce sont les finalités de ces épreuves quant à leurs résultats qui expliquent ces différences de perceptions des EE et non les croyances des enseignants, ni leur profil professionnel.

22. Effets de l'évaluation externe sur les pratiques des enseignants (QR2)

Nous synthétisons et interprétons dans ce point les résultats des deux parties empiriques qui ont trait à la deuxième question de recherche (QR2). Nous vérifions tout d'abord l'acceptation ou le rejet des hypothèses qui lui sont sous-jacentes, puis nous répondons à la question de recherche. Ces résultats sont discutés dans le chapitre 6 en les confrontant à la littérature.

QR2

Quels sont les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants du primaire ?

H2.1

L'évaluation externe des acquis des élèves affecte les pratiques des enseignants dans tout le processus d'enseignement (planification de l'enseignement, enseignement-apprentissage, évaluation de l'apprentissage).

Dans la première partie de notre recherche, cinq effets de l'EE sur les pratiques des enseignants ont été identifiés dans le discours des huit enseignants interrogés (tableau 10 ; p.149) :

- *Teaching to the test* ;
- Rétrécissement du curriculum enseigné ;
- Développement du professionnalisme des enseignants ;
- Encouragement à la collaboration ;
- Harmonisation des pratiques d'évaluation sommative.

Les analyses quantitatives confirment à plus large échelle que les enseignants reconnaissent ces effets dans le contexte de la recherche. Premièrement, les analyses de validation du questionnaire montrent que les enseignants regroupent les items définis *a priori* dans les dimensions attendues pour chaque effet. Deuxièmement, aucun résultat ne montre d'absence d'effet (l'effet de rétrécissement du curriculum montre toutefois un effet plancher pour l'EE de référence ($m=1.94$)).

L'hypothèse est confirmée. Les effets de l'évaluation externe des acquis des élèves touchent différents domaines de pratiques s'observent dans tout le processus d'enseignement. D'une part, ils touchent des "gestes techniques" de l'enseignant : la planification des apprentissages (rétrécissement du curriculum enseigné), l'enseignement des contenus (*teaching to the test*) et l'évaluation des apprentissages (harmonisation des pratiques d'évaluation sommative). D'autres effets sont plus transversaux et

concernent l'attitude de l'enseignant envers ses pratiques (développement du professionnalisme des enseignants, encouragement à la collaboration).

H2.2

L'EE d'orientation produit des effets plus grands sur les pratiques des enseignants que l'EE de référence.

Seule la partie quantitative de ce travail permet de vérifier cette hypothèse. Premièrement, la partie exploratoire n'investigue que des enseignants confrontés à l'épreuve d'orientation⁹¹. Deuxièmement, les analyses statistiques permettent de quantifier les effets et donc de les comparer entre les deux types d'EE.

Les résultats issus du questionnaire (tableau 22 ; p.175) renseignent tout d'abord sur la grandeur des effets créés par ces deux types d'épreuve externe sur les pratiques des enseignants⁹².

- *EE de référence.* Pour ce type d'EE, les effets mesurés sur les pratiques des enseignants sont faibles, à l'exception de l'harmonisation des pratiques d'évaluation sommative. Elle amène faiblement les enseignants à des pratiques de *teaching to the test* (m=2.11) ou de rétrécissement du curriculum (m=1.94). Elle les amène également faiblement à développer leur professionnalisme (m=2.42) et à travailler davantage en équipe (m=2.65). Par contre, l'effet d'harmonisation de leur pratiques d'évaluation sommative est modéré (m = 3.40).
- *EE d'orientation.* Pour ce type d'EE, les effets mesurés sur les pratiques des enseignants sont modérés. Elle amène modérément les enseignants à des pratiques de *teaching to the test* (m=4.02), de rétrécissement du curriculum (m=3.34), à harmoniser leur pratiques d'évaluation sommative (m=3.36) et à travailler davantage en équipe (m=3.24). Ce type d'EE amène faiblement à modérément les enseignants à développer le professionnalisme (m=2.96).

Les analyses statistiques permettent de comparer la grandeur des effets entre les deux types d'EE. Les résultats montrent que l'EE d'orientation entraîne significativement davantage d'effets que l'EE de référence (figure 22 ; p.176). Les différences sont très importantes pour les effets de *teaching to the test* et de rétrécissement du curriculum enseigné. Elles sont également significatives pour les effets de développement du professionnalisme des enseignants et de l'encouragement à la collaboration.

⁹¹ Il aurait peut-être été pertinent d'utiliser une méthodologie qualitative pour interroger également des enseignants confrontés à une évaluation externe à visée formative. Toutefois, dans un but de mesure et de comparaison, la méthode quantitative a été directement engagée.

⁹² Nous considérons la grandeur des effets de la manière suivante, selon leur moyenne : < 3 = faible ; 3 à 4 = modéré ; > 4 = fort. Les items ont été construits dans ce but.

Néanmoins, l'effet d'harmonisation des pratiques d'évaluation sommative ne diffère pas significativement entre les deux types d'EE.

L'hypothèse est confirmée, à une exception près. Les résultats montrent que les effets de l'EE sur les pratiques des enseignants sont plus grands pour l'EE d'orientation, sauf celui d'harmonisation des pratiques d'évaluation sommative qui ne diffère pas significativement.

H2.3

Pour un même type d'évaluation externe des acquis des élèves, les enseignants ne s'accordent pas sur l'importance de ses effets sur leurs pratiques.

Dans la partie exploratoire de la thèse, des effets ont pu être relevés dans le discours de certains enseignants, mais pas dans d'autres (p.119-141). De même, certains enseignants relèvent de nombreux effets (quatre effets ont été identifiés dans le discours de Vincent, Chantal, Julien et Romain), d'autres très peu (un seul effet est identifié dans le discours de Ghislain).

Pour chaque type d'EE, les analyses quantitatives montrent une certaine dispersion dans les réponses des enseignants pour tous les effets (tableau 22 ; p.175). La variabilité des effets de *teaching to the test* et de rétrécissement est moins grande pour l'EE de référence.

L'hypothèse est donc confirmée. Pour un même type d'évaluation externe des acquis des élèves, les enseignants ne s'accordent pas sur l'importance des effets sur leurs pratiques. Ce constat montre qu'il est alors nécessaire de comprendre cette variabilité, pour chaque type d'épreuve.

H2.4

Des liens existent entre certains effets de l'évaluation externe des acquis des élèves.

Les analyses quantitatives de la deuxième partie empirique démontrent des corrélations (modérées à fortes) entre les effets de l'EE (tableau 23 ; p.177). Ces corrélations sont semblables pour les deux types d'épreuve (légèrement plus importantes pour l'EE d'orientation). Ces analyses montrent que tous les effets identifiés, à l'exception de l'harmonisation des pratiques d'évaluation sommative, s'attirent mais qu'aucuns ne se repoussent (les corrélations sont toutes positives). Ainsi, plus un enseignant perçoit un effet, plus il perçoit les autres.

Les corrélations les plus fortes sont observées entre les effets de *teaching to the test* et de rétrécissement du curriculum. Ce résultat atteste que les enseignants qui sont amenés à "entraîner" leurs élèves à la réussite de l'épreuve sont aussi ceux qui vont privilégier les contenus évalués. En

outre, ces deux effets montrent des corrélations positives avec tous les autres, sauf celui d'harmonisation des pratiques d'évaluation sommative (sans toutefois montrer de corrélation négative).

Il est à relever que l'effet de développement du professionnalisme des enseignants montre de fortes corrélations avec tous les autres, sans exception. Premièrement, le lien entre cet effet, l'encouragement à la collaboration et l'harmonisation des pratiques d'évaluation sommative paraît logique, étant aujourd'hui attendu des enseignants que, pour être des "professionnels compétents", ils soient capables de travailler en équipe et de réguler leurs pratiques. Deuxièmement, le lien de cet effet avec les pratiques de *teaching to the test* et de rétrécissement du curriculum est plus difficile à interpréter. Les enseignants qui mettent en place des pratiques pour la réussite de leurs élèves aux épreuves externes, sont également ceux qui veulent montrer qu'ils sont de bons professionnels. Les enjeux ressentis expliquent certainement ce lien.

L'hypothèse est donc confirmée, il existe des liens entre les effets de l'EE.

Synthèse QR2 : Quels sont les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants du degré primaire ?

L'analyse des données issues des entretiens identifie cinq effets dans le contexte de la recherche. Les analyses statistiques confirment l'existence de ces cinq effets sur un plus large échantillon, portant également sur deux types d'épreuve externe. Ces effets touchent les pratiques des enseignants dans un large spectre. Ils touchent des "gestes techniques" de l'enseignement qui amènent l'élève à apprendre (planification de l'apprentissage, enseignement des contenus, évaluation de l'apprentissage). Les effets touchent également l'attitude des enseignants envers leurs pratiques (professionnalisme, collaboration).

Les analyses statistiques démontrent que l'EE d'orientation entraîne des effets significativement plus forts que l'EE de référence, à l'exception de l'effet d'harmonisation des pratiques d'évaluation sommative qui n'est significativement pas différent d'un type d'EE à l'autre.

En outre, les analyses qualitatives et quantitatives montrent que, pour un même type d'EE, les enseignants ne s'accordent pas sur l'importance des effets sur leurs pratiques. Ce constat montre qu'il est alors nécessaire de comprendre cette variabilité, pour chaque type d'épreuve.

Enfin, les résultats montrent que des liens modérés à forts existent entre les différents effets, dans les deux types d'EE.

23. Facteurs qui influencent les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants (QR3)

Nous synthétisons et interprétons dans ce point les résultats des deux parties empiriques qui ont trait à la troisième question de recherche (QR3). Nous vérifions tout d'abord l'acceptation ou le rejet des hypothèses qui lui sont sous-jacentes, puis nous présentons une synthèse sous forme de réponse à la question de recherche. Ces résultats sont discutés dans le chapitre 6, en les confrontant à la littérature.

QR3

Quels sont les facteurs qui influencent les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants du primaire ?

Pour répondre à la question de recherche, l'influence de trois types de facteurs est observée. Ils sont schématisés dans la figure ci-dessous.

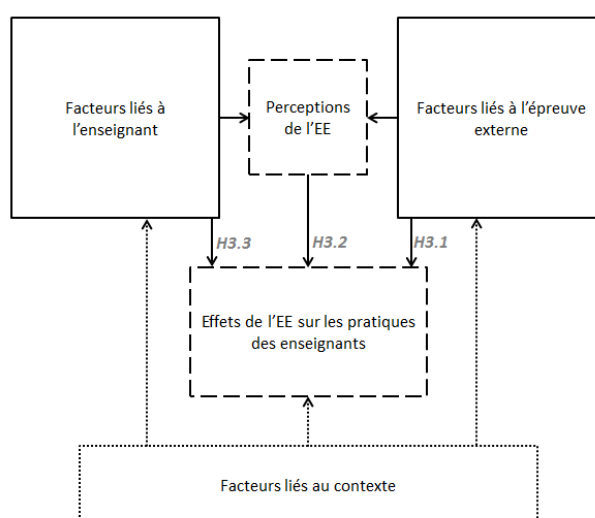


FIGURE 38 – Hypothèses sous-jacentes à QR3

H3.1

Les facteurs liés à l'épreuve externe influencent les effets de l'évaluation externe sur les pratiques des enseignants.

Les analyses quantitatives font apparaître des différences d'effet importantes entre les deux types d'EE (de référence ou d'orientation ; H2.2, p.210-211). Certaines caractéristiques de ces épreuves peuvent

expliquer ces différences⁹³. C'est surtout l'usage (interne et externe à la classe) de leurs résultats qui différencie ces deux types d'EE.

- *EE de référence*. Les résultats de ce type d'EE ne comportent pas d'enjeux élevés, ni pour l'élève, ni pour l'enseignant. Les résultats sont communiqués aux parents sous la forme d'une note (pas de comparaison à une norme possible pour les parents) et les résultats sont interprétés de façon critériée (par rapport à des objectifs du programme). La finalité des épreuves est claire pour les enseignants (elles portent le nom d'épreuve de "référence"). Elles sont parfois facultatives. Aussi, une seule branche est évaluée, le contenu à traiter n'est de ce fait pas trop étendu.
- *EE d'orientation*. Les résultats de ce type d'EE comporte des enjeux élevés pour la carrière des élèves. De cette épreuve dépend en partie leur orientation dans les filières de l'école secondaire. Ils comportent également des enjeux pour l'enseignant (l'évaluation externe confirme ou non son évaluation interne ; la comparaison des résultats internes et externes est observée par les cadres de la scolarité (inspecteurs, directeurs d'établissement secondaire) et par les parents. De plus, les résultats sont interprétés de façon normative (par rapport à une norme régionale). Pour les enseignants, cette épreuve est surtout vue comme une épreuve de sélection des élèves et non comme étant utile pour leurs pratiques. Elle ne se déroule pas en fin de cycle, mais en cours d'année ce qui entraîne la résistance de certains enseignants (trop tôt) et qui leur donne moins de temps pour traiter tous les contenus (les contenus sont pourtant sélectionnés et communiqués par les autorités). Plusieurs branches sont évaluées (contenu important).

L'hypothèse est confirmée. Les effets de l'évaluation externe sur les pratiques des enseignants sont inhérents aux caractéristiques de l'épreuve, mais surtout à l'usage de leurs résultats (de référence ou sommative). Le choix d'un type d'épreuve plutôt qu'un autre est donc la source d'effets spécifiques et de grandeur différente⁹⁴. Toutefois, la dispersion pour chaque type d'EE est grande. Il est donc nécessaire de voir quelles sont les facteurs qui influencent la variance des effets pour chaque type d'épreuve (H3.2, H3.3).

⁹³ Les épreuves sont décrites dans le point 14.2 (p.103-110).

⁹⁴ L'influence du type d'EE ne peut pas être quantifiée par une analyse de régression, les enseignants ayant répondu spécifiquement en référence à un type d'EE. Les analyses ont dès lors été séparées pour chaque type d'EE.

H3.2

Les facteurs liés aux perceptions des enseignants quant à l'épreuve externe influencent les effets de l'évaluation externe sur leurs pratiques.

Bien que se basant sur l'analyse d'un seul type d'EE, la première partie empirique apporte des premiers éléments de réponse. Dans le discours des enseignants interrogés, la plupart des variables qui influencent selon eux les effets de l'EE ont trait aux perceptions qu'ils ont de l'épreuve externe. Deux dimensions ont été identifiées, les enjeux en lien avec les résultats et l'estime du modèle de l'EE (tableau 10 ; p.149). Les analyses montrent que les enjeux influencent tous les effets identifiés. Le fait que les enseignants trouvent l'EE de qualité et la jugent comme utile pour leur pratiques les conduit à harmoniser leurs pratiques d'évaluation sommative.

Les analyses quantitatives permettent de mesurer l'influence de chaque perception quant aux effets de l'EE sur leurs pratiques par des analyses de corrélation (tableau 26 ; p.184) et de régression (tableau 36 ; p. 198). Certaines perceptions (enjeux et utilité) expliquent une grande part des effets suivants :

- *Enjeux ressentis face à l'EE.* La variable enjeux ressentis montre des corrélations positives et significatives avec les effets de *teaching to the test*, de rétrécissement du curriculum et de développement du professionnalisme. Par contre, aucun lien n'est trouvé avec l'encouragement à la collaboration et avec l'harmonisation des pratiques d'évaluation sommative. Ces corrélations sont modérées, sauf pour l'EE d'orientation où les liens sont forts avec les effets de *teaching to the test* et de rétrécissement du curriculum. Ces résultats sont confirmés par les analyses de régression qui montrent que les enjeux sont un prédicteur de poids pour ces trois effets, mais surtout pour les effets de *teaching to the test* et de rétrécissement du curriculum pour l'EE d'orientation.
- *Utilité de l'EE.* Pour les deux types d'EE, plus les enseignants perçoivent l'EE comme un moyen utile pour leurs pratiques, plus ils harmonisent leurs pratiques d'évaluation sommative interne. Cette corrélation est forte. De plus, une corrélation est également observée avec l'effet de développement du professionnalisme des enseignants (n'est pas significative pour l'EE d'orientation). Pour les deux types d'EE, les analyses de régression montrent que l'utilité perçue de l'EE est un prédicteur de poids pour les effets d'harmonisation des pratiques d'évaluation sommative et pour le développement du professionnalisme des enseignants.
- *Qualité de l'EE.* Pour l'EE de référence uniquement, plus les enseignants reconnaissent la qualité de l'EE, plus ils harmonisent leurs pratiques d'évaluation. Néanmoins, elle n'apparaît pas comme un prédicteur significativement important.
- *Concordance des pratiques d'évaluation interne et externe.* Le fait que les pratiques évaluatives des enseignants et celles de l'EE concordent a peu d'impact sur les effets de l'EE.

Cette perception de l'EE apparaît tout de même comme un prédicteur significatif pour le développement du professionnalisme des enseignants (EE d'orientation uniquement). Il est toutefois négatif, signifiant que plus les pratiques concordent, moins l'EE amène l'enseignant à développer son professionnalisme.

L'hypothèse est confirmée. Les effets de l'EE sur les pratiques des enseignants dépendent des perceptions des enseignants quant à l'épreuve. Plus les enjeux ressentis par les enseignants face à l'EE sont importants, plus ils en viennent à des pratiques de *teaching to the test* et de rétrécissement du curriculum, surtout en cas d'EE d'orientation. Ils amènent aussi les enseignants à développer leur professionnalisme. Leur perception de l'utilité de l'EE les incite à harmoniser leurs pratiques d'évaluation sommative. Le développement du professionnalisme des enseignants et l'encouragement à la collaboration (bien que faibles) sont expliqués conjointement par les enjeux et l'utilité des épreuves. La qualité des épreuves et la concordance des pratiques évaluatives internes et externes n'ont pas d'impact significatif.

H3.3

Les facteurs liés à l'enseignant influencent les effets de l'évaluation externe sur leurs pratiques (expérience professionnelle et croyances à propos de l'enseignement).

L'analyse des entretiens montre certaines tendances quant à l'influence de l'expérience professionnelle. Certains récits révèlent que l'expérience professionnelle a un impact sur les effets de l'EE. Ainsi, les enseignants expérimentés rapportent peu d'effets (Claude ; p.122 et Ghislain ; p. 140). Au contraire, pour Vincent (p.125), son manque d'expérience est la source de nombreux effets. Ce facteur l'amène à harmoniser ses pratiques d'évaluation sommative, à développer son professionnalisme, à travailler davantage en équipe et à introduire des pratiques de *teaching to the test*. Mais ce n'est pas le cas de Patrick (p.132), pour qui le manque d'expérience n'amène pas spécifiquement d'effet. Il est toutefois soutenu par des collègues plus expérimentés. Ces différences dues à l'expérience professionnelle peuvent s'expliquer par le fait que les enseignants expérimentés ont confiance en leurs pratiques, elles sont en quelque sorte déjà validées. De ce fait, l'expérience (notamment d'EE) permet aux enseignants d'asseoir une certaine confiance en leurs pratiques et d'atténuer les effets de l'EE (Aurélien, p.119). Les enseignants débutants ont par contre besoin d'un regard extérieur pour valider leurs pratiques.

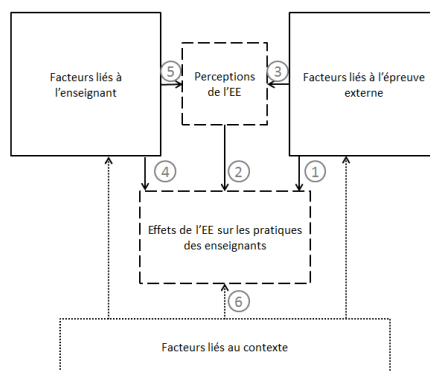
Les analyses quantitatives montrent certaines différences d'effets dues à l'expérience professionnelle. Certains effets de l'EE s'observent effectivement davantage pour les jeunes enseignants, d'autres pour les plus anciens. Ces différences dépendent du type d'EE (p.176-177). D'une part, l'EE de référence amène davantage les enseignants les plus expérimentés (plus de 30 ans d'expérience) à harmoniser

leurs pratiques d'évaluation sommative et à développer leur professionnalisme. L'expérience professionnelle est d'ailleurs un prédicteur significatif pour le développement du professionnalisme pour l'EE de référence (tableau 36 ; p. 198). Ces résultats montrent que les EE de référence servent davantage aux enseignants les plus expérimentés. Elle est pour eux davantage une opportunité de formation continue. D'autre part, l'EE d'orientation amène davantage les enseignants débutants à des pratiques de *teaching to the test* et à développer leur professionnalisme, craignant certainement davantage pour la validation de leurs pratiques. L'expérience n'est toutefois pas un prédicteur de poids. Il est à noter qu'aucune corrélation n'est observée entre l'expérience professionnelle et l'auto-efficacité perçue.

Pour chaque type d'EE, les croyances des enseignants à propos de l'enseignement ont un impact modéré sur les effets de l'EE (tableau 36 ; p. 198). Quelques différences sont observées (p.190-191). Premièrement, les enseignants qui se sentent moins efficaces affirment pratiquer davantage de rétrécissement du curriculum dans le cadre d'une EE de référence. Deuxièmement, pour ce type d'épreuve externe, les enseignants qui ne rejettent pas les conceptions behavioristes de l'enseignement sont amenés à davantage de rétrécissement du curriculum, de développement de leur professionnalisme, de collaboration et d'harmonisation de leurs pratiques d'évaluation sommative. Pour l'EE d'orientation, les enseignants qui ne rejettent pas la vision normative de l'évaluation sont amenés à pratiquer davantage de *teaching to the test*.

L'hypothèse est partiellement confirmée. Les facteurs liés à l'enseignant ont une influence modérée sur les effets de l'EE. Leur expérience professionnelle n'a que peu d'impact, si ce n'est en début et en fin de carrière. Les croyances des enseignants ont une influence modérée sur tous les effets. C'est surtout leurs conceptions de l'apprentissage qui créent des différences. Leurs conceptions de l'évaluation, et de leur auto-efficacité n'ont aucun impact significatif.

Synthèse QR3 : Quels sont les facteurs qui influencent les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants ?



(1) Des différences significatives pour tous les effets (sauf celui d'harmonisation des pratiques) sont observées entre les deux types d'EE (de référence ou d'orientation). Les différents types d'EE amènent des effets spécifiques. Toutefois, la dispersion des effets pour chaque type d'EE est grande. Il est alors nécessaire d'expliquer cette variance.

(2) Tout d'abord, les analyses montrent que ce sont surtout les enjeux et l'utilité de l'EE qui influencent ces effets. Les enjeux amènent davantage les enseignants à des pratiques de *teaching to the test* et de rétrécissement du curriculum. L'utilité les amène davantage à harmoniser leurs pratiques d'évaluation sommative. Le développement du professionnalisme des enseignants et l'encouragement à la collaboration (bien que faibles) sont expliqués conjointement par les enjeux et l'utilité des épreuves. La qualité des épreuves et la concordance des pratiques évaluatives internes et externes n'ont pas d'impact significatif sur les effets. (3) En outre, les perceptions des enjeux et de l'utilité de l'EE dépendent du type d'épreuve (H1.2).

(4) Ensuite, les analyses démontrent que les facteurs liés à l'enseignant contribuent faiblement à expliquer les effets de l'EE. Leur expérience professionnelle n'a que peu d'impact, si ce n'est en début et en fin de carrière. Les croyances des enseignants ont une influence modérée sur tous les effets. C'est surtout leurs conceptions de l'apprentissage qui créent des différences. Leurs conceptions de l'évaluation, et de leur auto-efficacité n'ont aucun impact significatif. (5) En outre, les facteurs liés aux enseignants n'influencent pas leurs perceptions (H1.3, H1.4).

(6) Les facteurs liés au contexte ne sont pas étudiés dans ce travail, bien qu'ayant certainement un impact sur les effets de l'EE (voir point 7.1 ; p.79).

En conclusion, les facteurs qui influencent significativement les effets de l'EE sont premièrement le choix d'un type d'épreuve, puis la perception par les enseignants des enjeux et de l'utilité de ces dispositifs (ces perceptions sont liées au type d'épreuve). Le choix d'un type d'EE (de référence ou d'orientation) détermine donc les effets de l'EE. Leur perception par les enseignants atténue ou augmente ces effets.

CHAPITRE 6

CONCLUSIONS

24. L'EE, un "mal nécessaire" pour les enseignants	220
25. Des effets bénéfiques et néfastes de l'EE	222
26. L'usage des résultats de l'EE comme prédicteur principal de ses effets	227
27. L'EE, un moyen de perfectionnement continu pour les enseignants	228
28. Limites et prolongements du travail	234

24. L'EE, un “mal nécessaire” pour les enseignants

Le premier objectif de la thèse est de comprendre comment les enseignants perçoivent l'évaluation externe des acquis des élèves (QR1, p.203-207). Les résultats des deux parties empiriques de notre recherche, grâce à des méthodes différentes mais complémentaires, confirment un constat déjà présent dans la littérature spécialisée et observé tant dans les contextes européens que nord-américains (Dierendonck, 2008; B. Jones, 2007; Lafontaine et al., 2009; Pons, 2011). Les enseignants perçoivent le plus souvent l'évaluation externe de manière contradictoire, comme “mal nécessaire”⁹⁵. En effet, l'évaluation externe est bien acceptée par les enseignants, mais d'autre part elle fait l'objet de résistances parfois vives. Nous exposons ce paradoxe en confrontant nos résultats à la littérature.

24.1. Un instrument nécessaire

L'évaluation externe est de manière générale reconnue par les enseignants comme une mesure de qualité (garante d'objectivité et digne de confiance), parfois même davantage que leur propre évaluation car les épreuves sont créées par des spécialistes, sont testées avant leur passation, etc. Les enseignants sont alors plutôt favorables à de tels outils (Lafontaine et al., 2009). A l'image d'autres enquêtes (Cattonar et al., 2010; Dierendonck, 2008; Laveault, 2009b; Longchamp, 2010; Soussi et al., 2006) notre recherche démontre que les épreuves externes fribourgeoises sont considérées par la plupart des enseignants comme un moyen de se rassurer (elles permettent de situer la performance de leurs élèves en référence à une norme), comme un modèle à suivre (elles montrent les “bonnes pratiques”) et comme un guide (elles donnent des balises concrètes sur les attentes des programmes). Notre étude confirme que, quel que soit le type d'épreuve externe (de référence ou d'orientation), les enseignants s'accordent sur sa qualité. Néanmoins, alors que les deux types de dispositifs ont notamment comme objectif de donner des repères aux enseignants, l'évaluation externe de référence est perçue comme davantage utile pour leurs pratiques (figure 25 ; p.182).

De plus, notre étude montre que, pour un même type d'EE, l'importance de ce sentiment d'utilité de l'EE varie entre les enseignants. Selon Coburn (2004), l'appropriation de l'évaluation externe par les enseignants dépend de plusieurs facteurs liés à l'épreuve. Notre étude illustre tout à fait ce constat :

- *Degré d'opportunité.* Les enseignants doivent trouver en l'EE des opportunités de développement de leurs pratiques. Notre étude démontre que les enseignants ont généralement le sentiment d'y trouver une utilité pour leurs propres pratiques (tableau 24 ; p.181). Nos

⁹⁵ Cette expression est utilisée par certains enseignants dans les entretiens.

résultats montrent que cette utilité (donc cette opportunité) dépend notamment de la qualité de l'EE (tableau 25 ; p.183) ;

- *Degré d'importance.* Les opportunités de développement offertes par l'EE doivent être importantes. Notre étude démontre que l'impression d'utilité de l'EE des enseignants dépend surtout des finalités de l'EE. L'EE de référence est perçue comme davantage utile par les enseignants que l'EE d'orientation (figure 25 ; p.182) ;
- *Degré de congruence.* L'EE doit avoir une certaine congruence avec les pratiques du terrain. Notre étude montre que plus les enseignants reconnaissent leurs pratiques dans l'EE, plus ils perçoivent l'épreuve comme utile pour leurs pratiques (tableau 25 ; p.183) ;
- *Degré de volonté.* Pour être acceptés, les changements dus à l'épreuve externe ne devraient pas être imposés (normatif vs. régulateur), ni comprendre de sanctions. Notre recherche montre que les enseignants voient moins d'utilité pour leurs pratiques, lorsque les résultats de l'EE comportent des enjeux élevés (EE d'orientation) (figure 25 ; p.182).

24.2. Un instrument qui fait l'objet de résistances

Comme exposé dans la littérature, c'est surtout l'usage des résultats de l'EE qui fait l'objet de résistances de la part des enseignants. Les résultats des élèves peuvent en effet être la source de pression, les enseignants étant aujourd'hui tenus responsables des résultats des élèves. Les enseignants craignent donc que les résultats de ces épreuves soient une mesure de leurs compétences et/ou un moyen de pression. Les résultats sont généralement communiqués et utilisés de manière interne (par les autorités scolaires, par les enseignants). Cette pression est exacerbée lorsqu'à un usage des résultats interne à la classe (par l'enseignant) on ajoute un usage externe (par les autorités, les parents, la société). L'usage des résultats dépend des modes de régulation des systèmes (*accountability* dure ou douce)⁹⁶, mais aussi des caractéristiques des épreuves (Maroy, 2014; Mons & Pons, 2006; Pons, 2011). Notre étude confirme que les enseignants ressentent, à des degrés variables, une certaine pression liée à l'usage des résultats de l'EE. Elle dépend des finalités de l'EE (de référence ou d'orientation). L'épreuve d'orientation est perçue comme stressante (les résultats conditionnent la carrière de l'élève et valident les pratiques des enseignants), bien davantage que l'évaluation externe de référence (les résultats ont un but d'information). Néanmoins, pour cette dernière, l'enseignant ressent tout de même une pression modérée, comparant lui-même la performance de sa classe à une norme.

⁹⁶ Ces différents modes de régulation sont l'objet du point 3.4.4 (p.20-23).

25. Des effets bénéfiques et néfastes de l'EE

Notre recherche a permis d'investiguer les effets qu'entraîne l'évaluation externe des acquis des élèves sur les pratiques des enseignants (QR2, p.209-212). Cinq effets ont été identifiés dans le contexte de la recherche⁹⁷. Ils s'observent sur l'ensemble du processus de l'enseignement. Trois effets touchent des "gestes didactiques" de l'enseignant (*teaching to the test*, rétrécissement du curriculum enseigné, harmonisation des pratiques d'évaluation sommative interne) ; les deux autres touchent des attitudes de l'enseignant envers ses pratiques (développement du professionnalisme des enseignants, encouragement au travail en équipe). Ces effets ont été identifiés dans d'autres contextes. Dans la littérature scientifique, les effets sont considérés soit comme attendus ou inattendus, soit comme bénéfiques ou indésirables, voire néfastes. Notre but ici est de discuter les effets identifiés, notamment en confrontant nos résultats à la littérature.

25.1. Des effets néfastes pour les élèves, les enseignants et le système

Notre étude identifie, dans le contexte de la recherche, deux effets reconnus par la littérature comme indésirables, inattendus, voire néfastes. Il s'agit de l'effet de *teaching to the test* et de rétrécissement du curriculum enseigné⁹⁸. L'EE amène les enseignants interrogés à adopter des pratiques d'entraînement intensif, que l'on peut définir comme du *coaching*, pour préparer leurs élèves à réussir le test plutôt qu'à développer les apprentissages inscrits dans les plans d'études (*teaching to the test*) (Darling-Hammond, 1994 ; Laveault, 2009b). Elle peut aussi amener l'enseignant à donner la préférence aux branches et/ou à des compétences plus basiques qui seront évaluées dans l'épreuve externe (rétrécissement du curriculum enseigné) (Monseur & Demeuse, 2005). Ces deux effets vont généralement de pair (Au, 2007), ce qui est confirmé dans notre étude par une corrélation importante (tableau 23 ; p.177).

Ces deux effets sont néfastes pour les différents acteurs de la scolarité :

⁹⁷ Certains effets de l'EE sur les pratiques des enseignants reconnus par la littérature (p.54-72) n'ont pas été identifiés dans le discours des enseignants interrogés : alignement des contenus enseignés, focalisation de leur attention sur certains types d'élèves, tricheries. Aussi, si les enseignants ont relevé l'harmonisation de leurs pratiques d'évaluation sommative, ils n'ont pas fait cas d'un tel effet sur leurs pratiques d'enseignement-apprentissage. Enfin, les enseignants n'ont pas relevé non plus d'impact des épreuves externes sur leurs pratiques d'évaluation formative.

⁹⁸ D'autres effets néfastes apparaissent également dans la littérature (tels que la démotivation ou la déprofessionnalisation des enseignants, des tricheries), mais n'ont pas été identifiés dans le contexte étudié.

- *Elèves.* Bien que ces pratiques puissent améliorer la réussite de certains élèves à court terme, elles ne leur permettent pas de réaliser un véritable apprentissage à long terme des objectifs inscrits dans les programmes. Elles avantagent les élèves les plus compétents. Certains apprentissages manquent aussi aux élèves dans leur cursus scolaire et certaines de leurs compétences sont également préteritiées (créativité, autonomie, etc.).
- *Enseignants.* Alors que l'EE peut être le vecteur de changements positifs dans les pratiques des enseignants, elle peut aussi les amener vers des pratiques qui vont à l'encontre de leurs conceptions et/ou à l'encontre des théories de l'enseignement et de l'apprentissage.
- *Système.* Ces effets ont également des répercussions négatives sur le système. Elles renforcent les disparités entre élèves. De plus, elles ne permettent pas l'alignement des contenus entre les années et le respect des plans d'études. Enfin, à cause de ces pratiques de préparation, les résultats des élèves ne reflètent pas tout à fait leur niveau d'apprentissage réel. Ceci ne permet ni aux élèves, ni à l'enseignant, ni aux autorités de recenser des informations fiables.

En outre, notre étude démontre que les effets de *teaching to the test* et de rétrécissement du curriculum apparaissent surtout dans le cas de l'épreuve d'orientation, très peu pour les épreuves de référence (figure 22 ; p.176). Ce sont surtout des enjeux élevés ressentis par l'enseignant face à l'EE qui expliquent ces effets (tableau 36 ; p.198). Ceci confirme certains résultats de recherche (M. Gail Jones et al., 2003; Madaus, 1991).

25.2. Des effets bénéfiques pour les élèves, les enseignants et le système

Il est généralement accepté par la littérature que l'évaluation externe des acquis des élèves peut avoir des effets bénéfiques pour les élèves (et leurs parents), les enseignants et le système (Crahay, 2009; Monseur & Demeuse, 2005). Elle est une source d'informations importantes et utiles pour chaque acteur. Premièrement, elle assure de bénéficier d'une évaluation "équitable et objective" durant le cursus scolaire (surtout dans un but de certification et/ou d'orientation). Deuxièmement, elle permet de situer les performances individuelles et collectives par rapport à une norme (interprétation normative) et/ou par rapport à des objectifs des programmes (interprétation critériée). Troisièmement, elle permet d'agir sur les zones critiques (pilotage).

Il est notamment attendu par les autorités que les épreuves externes soient un moyen efficace pour orienter les pratiques des enseignants. Elles seraient même un vecteur bien plus efficace que les programmes d'études, les moyens d'enseignement ou la formation pour atteindre le terrain (Mons & Crahay, 2011). Ces ajustements peuvent intervenir sur différents domaines de la classe, enseignement (matériel didactique, modèle d'enseignement-apprentissage, planification de l'apprentissage, accompagnement des élèves, évaluation des apprentissages) (Spratt, 2005; Wall, 2000). Des études

montrent cependant que ce sont surtout sur les pratiques d'évaluation des enseignants qui sont le plus touchées par ce levier didactique (Cattonar et al., 2010; Dierendonck, 2008; Longchamp, 2010). L'EE fait évoluer leurs pratiques, leur donne de nouvelles idées (Carette, 2007; Laveault, 2009b). Notre étude identifie justement, l'effet d'harmonisation des pratiques d'évaluation sommative des enseignants interrogés. Dans le contexte de l'étude, l'évaluation externe est utilisée par une partie des enseignants comme un modèle pour évaluer les apprentissages des élèves à des fins sommatives (bilan en fin de séquence d'enseignement-apprentissage).

Toutefois, l'impact des EE sur les pratiques est interrogé par la recherche, il serait plutôt modéré. Des recherches issues de différents contextes montrent que les enseignants sont plutôt partagés (Carette & Dupriez, 2009; Lafontaine et al., 2009; Longchamp, 2010; Pons, 2011; Soussi et al., 2006). Une partie d'entre eux ne font pas usage de l'évaluation externe pour modifier leurs pratiques. Notre étude confirme ce constat, l'effet d'harmonisation observé sur les pratiques d'évaluation sommative est plutôt modéré (tableau 22 ; p.175).

Aussi, l'effet d'harmonisation des pratiques serait plus fort lorsque les enjeux de l'EE sont élevés (Popham, 1987). Notre recherche ne démontre pourtant aucun lien entre enjeux et harmonisation, sans montrer non plus le contraire (tableau 26 ; p.184). En outre, les enjeux ressentis par les enseignants ne sont pas un prédicteur de cet effet (tableau 36 ; p.198) et aucune différence significative n'est observée entre les différents types d'épreuve (de référence ou d'orientation) (figure 22 ; p.176). Dans certains contextes (*high stakes tests*), des enjeux élevés seraient même plutôt néfastes. Outre des pratiques de *teaching to the test* ou de rétrécissement du curriculum, ils peuvent aussi être le vecteur de mauvaises pratiques évaluatives (questionnaires à choix multiples, interprétation normative, niveau bas de compétences testées, etc.) (Au, 2007; Darling-Hammond & Wise, 1985).

La recherche s'interroge également sur les résultats effectifs de ce levier didactique dans les classes. Les changements de pratique promus par l'EE seraient superficiels, les enseignants donnant plutôt un message de complaisance (Carette & Dupriez, 2009; Lafontaine et al., 2009). En effet, le « couplage » entre pratiques attendues par les autorités et celles effectives dans les classes peut parfois être manqué (Dumay et al., 2013). L'implantation des pratiques attendues sur le terrain peut être très variable, allant du rejet à l'accommodation (Coburn, 2004). Nos résultats illustrent bien ce constat, deux types d'implantation ayant été observés dans les entretiens (p.144) :

- *Intégration du modèle.* Certains enseignants, convaincus par le modèle évaluatif des épreuves externes, modifient leurs pratiques (assimilation ou accommodation selon Coburn) ;
- *Imitation du modèle.* Certains enseignants, ne reconnaissant pas le modèle, utilisent les pratiques attendues sans pourtant renoncer à leurs pratiques pré-existantes (découplage – réponse symbolique ou structure parallèles selon Coburn) ;

- *Rejet.* Certains enseignants rejettent totalement le modèle (rejet selon Coburn). Toutefois, notre étude démontre peu de rejet (aucun enseignant durant les entretiens et très peu selon le questionnaire).

Ce n'est donc pas simplement par des injonctions politiques (ici des épreuves externes) que des changements de pratiques peuvent être réalisés sur le terrain. Selon la littérature, le changement n'est possible que lorsqu'il est progressif, accessible et n'entraîne pas de résistances (Carette & Dupriez, 2009). L'implantation de nouvelles pratiques dépend également de leur alignement, de leur cohésion externe et interne et de leur intégration dans un cadre général (Laveault, 2012b). Notre étude du contexte fribourgeois illustre bien ce constat général. Comme démontré auparavant, l'effet d'harmonisation dépend de l'appropriation de l'épreuve par les enseignants. Ceci est illustré par nos résultats qui montrent des liens positifs entre l'effet d'harmonisation et l'utilité, la qualité et la concordance des pratiques (tableau 26 ; p.184). L'utilité de l'EE perçue par les enseignants est d'ailleurs un prédicteur important de cet effet (tableau 36 ; p.198).

En ce qui concerne les EE fribourgeoises, l'effet modéré d'harmonisation des pratiques peut s'expliquer par la faiblesse ou la quasi absence de feedback fourni aux enseignants, aussi bien pour les EE de référence que celles d'orientation. Dans les deux cas, les enseignants ne peuvent qu'en faire une interprétation normative (comparaison des résultats à une norme cantonale). En ce sens, les épreuves cantonales fribourgeoises ne permettent pas une véritable « régulation » de leurs pratiques (Allal, 2007; Laveault, 2007; Sadler, 1989)⁹⁹. La « boucle » d'un état initial à un état final souhaité ne comprend aucune rétroaction. Outre un feedback peu effectif, aucune action de remédiation n'est comprise dans le processus, ou celle-ci reste à l'initiative personnelle de l'enseignant. Aucun accompagnement spécifique n'est prévu pour aider les enseignants à atteindre l'objectif (développer-ajuster ses pratiques). Par ce manque de rétroaction, les informations risquent d'être mal interprétées ou tout simplement peu ou pas exploitées par les enseignants. Pour nous, un meilleur feedback (interprétation critériée des résultats par objectif ou par item, par exemple) et un accompagnement (individuel ou collectif) pourrait augmenter sensiblement l'utilité des épreuves externes aux yeux des enseignants et permettre le développement ou l'ajustement de leurs pratiques d'enseignement et d'évaluation (cet argument est discuté dans le point 27, p.228-233).

25.3. Des effets attendus qui restent attendus

Comme le montre la littérature, il est généralement attendu des autorités que l'EE amène les enseignants, rendus aujourd'hui responsables des résultats de leurs élèves, à travailler davantage et à

⁹⁹ Le terme de « régulation » est discuté au point 5.2.2 (p.36-37).

améliorer leur professionnalisme (Klieme et al., 2004; Maroy & Cattonar, 2002). On attend aujourd'hui de l'enseignant qu'il soit un spécialiste de l'apprentissage, un praticien réflexif et qu'il travaille en équipe. Pourtant cet effet est mis en doute, il resterait de « nature instrumentale », l'EE n'engageant pas les enseignants dans une voie de développement de leur professionnalisme, surtout lorsqu'elle est de nature réflexive et que les enjeux sont faibles (Mons & Dupriez, 2010).

Notre étude montre effectivement que cet effet attendu ne se réalise que peu dans le contexte fribourgeois. L'EE amène faiblement les enseignants à davantage de réflexion et de rigueur sur leurs pratiques. Elle les amène modérément à travailler davantage en équipe (tableau 22 ; p.175). De plus, les analyses de corrélation (tableau 26 ; p.174) et de régression (tableau 36 ; p.198) montrent que cet effet est lié aux enjeux et qu'ils sont plus importants pour l'EE d'orientation (figure 22 ; p.176). Toutefois, les résultats démontrent également que ces effets sont aussi, et dans des proportions semblables, liés à l'utilité perçue des épreuves (tableau 22 ; p.175). En outre, notre étude démontre aussi qu'ils dépendent de l'expérience des enseignants, étant surtout observés chez les enseignants les plus expérimentés (p.178-179).

La faible importance de l'effet de développement du professionnalisme et d'encouragement à la collaboration (pour les deux types d'EE) montre que soit les enseignants n'ont pas besoin de ce genre d'épreuve pour se montrer réflexifs et rigoureux dans leurs pratiques, soit que ces épreuves ne le permettent pas. Ce constat rejoint notre commentaire du point précédent concernant le manque de rétroaction offert par les deux types d'EE (peu de feedback, aucune remédiation)¹⁰⁰. Deuxièmement, pour que l'EE développe le professionnalisme des enseignants, il est préférable de promouvoir son utilité plutôt que ses enjeux, afin notamment d'éviter les effets néfastes de l'EE sur leur pratiques (*teaching to the test*, rétrécissement du curriculum), mais aussi sur leur attitude (déprofessionnalisation et démotivation).

¹⁰⁰ Le manque de rétroaction des EE fribourgeoises est discuté au point 27.1 (p.229-231).

26. L'usage des résultats de l'EE comme prédicteur principal de ses effets

L'un des objectifs de la thèse est d'identifier les facteurs qui influencent les effets de l'évaluation externe sur les pratiques des enseignants (QR3, p.212-217). Les résultats montrent que ce sont principalement les facteurs liés aux épreuves et à leurs perceptions par les enseignants qui influencent les effets de l'EE sur leurs pratiques (tableau 36 ; p.198). Le but de cette partie est de discuter l'influence de ces facteurs, notamment en confrontant nos résultats à la littérature.

Si les effets (bénéfiques ou néfastes) de l'EE sont relevés dans les recherches, rares sont celles qui tentent d'expliquer par quels processus ils se créent. Il semble pourtant important de déterminer la nature de ces influences (Carette & Dupriez, 2009). Dans le champ de la linguistique, certaines études portant sur le *backwash effect* s'interrogent sur ce qui provoque les effets de l'EE sur les pratiques des enseignants (Mizutani, 2009; Spratt, 2005; Wall, 2000). Les effets de l'EE sont une mécanique complexe qui inclut tous les acteurs de la scolarité à des niveaux de contexte différents. Des recherches se proposent de modéliser ces effets et les facteurs qui ont un impact sur ceux-ci (p.76-78). Elles relèvent d'ailleurs l'importance des croyances des enseignants à propos de l'enseignement. Néanmoins, ces modèles énumèrent les facteurs d'influence, sans pour autant les quantifier. Il s'agit de facteurs liés au contexte éducatif, à l'épreuve externe ou aux enseignants. Sur ces bases théoriques, nous avons également proposé un modèle a priori (p.88) que nous avons pu tester. Suite à nos analyses, nous proposons ici un modèle a posteriori (figure 38) que nous commentons.

Notre recherche démontre que les effets de l'EE sur les pratiques des enseignants fribourgeois dépendent surtout de l'usage de ses résultats. Les effets observés sont premièrement dus au type d'EE (figure 24 ; p.174). Si l'objet technique fait consensus (sans différence entre les types d'épreuves), ce sont les finalités quant à ses résultats (référence vs. orientation) qui déterminent ses effets. Deuxièmement, les perceptions qu'ont les enseignants de ces dispositifs expliquent la variabilité de ces résultats pour chaque type d'EE (tableau 36 ; p.198). Ces perceptions sont une sorte de filtre entre les caractéristiques des épreuves et leurs effets. Les perceptions qui sont porteuses de plus de signification sont également liées à l'usage (interne et/ou externe à la classe) des résultats de l'EE. La variabilité des effets de l'évaluation externe s'explique principalement par les enjeux ressentis (liés à l'usage interne et externe des résultats) et l'utilité perçue (usage interne des résultats). Finalement, les facteurs propres aux enseignants ont une influence très faible (tableau 36 ; p.198) et n'apparaissent pas dans notre modèle.

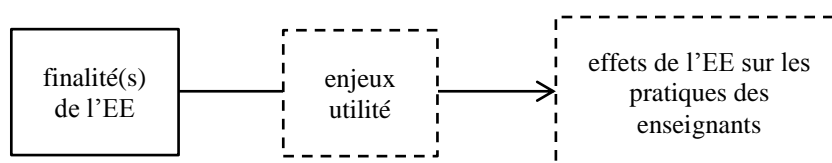


FIGURE 38 – Modélisation a posteriori des effets de l'EE et des facteurs qui les influencent

27. L'EE, un moyen de perfectionnement continu pour les enseignants

Les différents résultats de la thèse, les réponses aux différentes questions de recherche (QR1, QR2, QR3) mais aussi les apports théoriques, permettent de nous positionner quant à l'usage de l'EE, du point de vue des enseignants. D'après l'analyse du regard des enseignants fribourgeois, l'EE peut être un moyen de perfectionnement continu de leurs pratiques, toutefois selon certaines conditions. Nous développons ce constat dans ce point.

Lors des dernières décennies, l'évaluation externe des acquis des élèves a connu un développement important dans les systèmes éducatifs. Le recours à ce genre de dispositifs s'est accéléré au plan international, national et régional (Eurydice, 2009; Mons & Crahay, 2011; Pons, 2011). L'évaluation externe est aujourd'hui à la fois un instrument pédagogique et politique. Elle sert d'une part à garantir une évaluation objective et équitable des apprentissages des élèves (surtout dans un but de certification et de sélection équitable des élèves) et d'autre part elle sert au pilotage du système (observation du système et intervention sur le terrain). Ainsi, l'usage de l'évaluation externe doit permettre d'améliorer la qualité des systèmes éducatifs, en étant utile à la fois aux élèves, aux enseignants et aux autorités.

Cependant, le bienfondé de l'usage de l'évaluation externe des acquis des élèves dans un but d'améliorer la qualité des systèmes éducatifs (meilleure performance des élèves et des enseignants) n'est pas encore prouvé et reste débattu. En effet, comme démontré dans notre cadre théorique, la littérature présente des résultats bien souvent contradictoires. Davantage de résultats empiriques sont nécessaires pour comprendre sous quelle(s) forme(s) l'évaluation externe peut être une "valeur ajoutée" pour les systèmes éducatifs (et finalement pour la société en général). En effet, certaines conditions sont à respecter pour qu'elle soit porteuse de bénéfices et afin d'éviter ses effets néfastes. Même si aucun modèle idéal n'existe, des recommandations théoriques et pratiques sont faites par de nombreux auteurs quant à un usage approprié de l'évaluation externe dans les systèmes scolaires (voir point 9, p.85-87).

Pour les enseignants, l'évaluation externe des acquis des élèves se trouve généralement dans une tension entre contrôle et formation. Comme démontré précédemment (p.220), elle est un « mal nécessaire » (enjeux et utilité). Apparaît alors un paradoxe quant à l'usage de l'évaluation externe pour les pratiques enseignantes. D'un côté, il est attendu qu'elle permette à l'enseignant (praticien réflexif autonome) d'améliorer ses pratiques à la lumière des résultats. Pourtant, en parallèle, l'évaluation externe contraint cette autonomie. L'enseignant (ses compétences, son professionnalisme) est contrôlé « à distance » par l'évaluation externe, sous une forme nouvelle de responsabilisation liée aux résultats des élèves. Ce contrôle peut d'ailleurs amener à des résistances des enseignants et à des effets néfastes (Maroy, 2014; Mons, 2008a).

L'EE a acquis, dans de nombreux contextes éducatifs, une place privilégiée. Cependant, il est nécessaire d'agir sur les caractéristiques des dispositifs en place afin que chaque acteur du système éducatif puisse profiter de leurs effets bénéfiques et éviter leurs effets néfastes. En ce qui concerne les enseignants, il est nécessaire d'optimiser leur impact dans les classes. Tout en reconnaissant les limites et le contexte retreint de notre étude¹⁰¹, les résultats de nos analyses du regard des enseignants permettent d'illustrer et de développer certaines recommandations concernant l'usage de l'EE. En effet, comme démontré dans le contexte fribourgeois, les effets de l'EE sur leurs pratiques dépendent principalement des caractéristiques des épreuves externes (surtout l'usage de leurs résultats) et de leurs perceptions par les enseignants (surtout leurs enjeux et leur utilité) (tableau 36 ; p.198). En outre, notre étude démontrant qu'aucune corrélation n'existe entre les facteurs d'enjeux et d'utilité (tableau 25, p.183), il est nécessaire d'agir sur ces deux facteurs simultanément. Une intervention sur l'un n'aura pas forcément d'impact sur l'autre. Selon nous, pour que l'évaluation externe soit un "moyen de perfectionnement continu" pour les enseignants, il est nécessaire (1) d'optimiser son rôle de formation et de (2) clarifier/minimiser son rôle de contrôle. Finalement, il est également nécessaire d' (3) éviter qu'une même épreuve suive plusieurs finalités à la fois.

27.1. Optimiser le rôle de formation de l'EE pour les enseignants

Pour que l'évaluation externe puisse servir à tous les niveaux du système mais surtout aux enseignants, elle doit, selon nous, être non seulement une mesure externe des performances mais aussi un moyen de « régulation » des pratiques.

L'évaluation externe des acquis des élèves est une mesure individuelle ou collective de qualité tout à fait pertinente. Elle permet de situer la performance d'un élève, d'une classe, d'un établissement ou d'un système scolaire. Cependant, l'évaluation externe, pour servir aux différents acteurs du système et pour éviter les effets néfastes de son usage, ne doit pas être une simple mesure. En effet, la littérature scientifique critique l'aspect uniquement quantitatif de ces politiques de *testing* (Demailly, 2001; Duru-Bellat & Jarousse, 2001). Elles sont difficilement réalisables au niveau méthodologique, leurs résultats sont trop uniformisés (ne permettent pas de prendre en compte des situations particulières) et l'interprétation parfois trop difficile de ces résultats peut être le vecteur d'effets néfastes. En outre, le recours aux résultats des élèves comme unique indicateur relève de la commodité (mise en place rapide, moindre coût). De plus, ces données ne sont que peu exploitées, ni par les autorités ni par les enseignants. Le recours de plus en plus fréquent à l'évaluation externe dans un but de pilotage crée de véritables "cimetières de données", des quantités de résultats qui ne sont pas

¹⁰¹ Les limites de notre recherche sont relevées dans le point 28 (p.234).

exploitées. Certains auteurs mettent en doute que les enseignants puissent « lire » les résultats de l'évaluation externe dans le but de développer leurs pratiques (Carette & Dupriez, 2009). Les enseignants peuvent se retrouver démunis face aux résultats d'épreuves externes, ne sachant comment les interpréter ou n'en ayant pas les moyens. Et bien souvent, comme dans le contexte fribourgeois, cette interprétation est le plus souvent laissée à leur propre initiative. Selon les résultats de notre étude, si une épreuve externe offre des opportunités de développement des pratiques des enseignants (que cette opportunité est perçue par les enseignants), elle pourra davantage entraîner des effets positifs sur le terrain (alignement des contenus, harmonisation des pratiques, réflexivité, collaboration).

Pour que l'évaluation externe soit utile aux enseignants, pour qu'elles leur permettent de développer leurs compétences et leurs attitudes, ceux-ci doivent pouvoir bénéficier des conditions d'une véritable « régulation » (Allal, 2007; Laveault, 2007)¹⁰². Ce processus de perfectionnement prend l'image d'une boucle en quatre étapes :

1. L'objectif (les finalités) de l'épreuve externe doit être clair et transparent ;
2. L'épreuve externe doit donner aux enseignants des résultats détaillés de l'atteinte ou non de l'objectif (une mesure initiale, un diagnostic) ;
3. Les résultats doivent être suivis ou accompagnés d'un feedback et de moyens de remédiation pour améliorer cette performance initiale ;
4. L'atteinte de l'objectif doit pouvoir être confirmé ou non et/ou être redéfini.

En suivant ces quatre étapes, différents moyens sont envisageables afin de garantir et d'optimiser la régulation des pratiques des enseignants :

- *Expliciter les finalités de l'épreuve.* Les finalités de l'épreuve devraient être communiquées et clairement explicitées (voire discutées) aux différents acteurs (enseignants, élèves, parents) par les administrateurs de l'épreuve¹⁰³.
- *Fournir des résultats transférables dans la pratique.* Les résultats de l'épreuve devraient être suffisamment détaillés pour être exploités par les enseignants et/ou les autorités. Une interprétation critériée (par rapport à des objectifs et non à une norme) et détaillée permet d'observer plus spécifiquement les réussites et les difficultés des élèves. Des pistes didactiques doivent également être fournies aux enseignants. Aussi, les résultats devraient permettre aux enseignants d'observer l'évolution des performances. Laveault (2009b) propose d'ailleurs l'évaluation « intérimaire » (voir p.48).

¹⁰² La notion de régulation est discutée et définie au point 5.2.2 (p.36-37).

¹⁰³ Nous verrons par la suite, qu'il est préférable qu'une épreuve suive une finalité unique (point 27.3 ; p.232-234).

- *Soutenir les enseignants.* Les enseignants devraient être soutenus dans la compréhension et l'utilisation des épreuves externes, notamment de leurs résultats. Un enseignant ne devrait pas se retrouver seul face aux résultats de l'évaluation externe. Il devrait être soutenu en travaillant en équipe avec ses pairs (dans la cadre de l'établissement notamment) mais aussi avec le soutien des cadres intermédiaires (directeur d'établissement, inspecteur, responsable pédagogique...).
- *Intégrer les enseignants dans le processus d'évaluation externe.* Pour une meilleure appropriation de l'outil, les enseignants devraient être intégrés au processus de conception et d'administration de l'épreuve (conception, passation, correction, analyse des données).
- *Garantir la qualité et la validité de l'épreuve.* L'épreuve devrait être exemplaire au niveau méthodologique et scientifique, en lien avec les théories de l'apprentissage, de l'enseignement et de l'évaluation. Il serait également nécessaire qu'elle soit équitable (sa standardisation en est une condition). Le traitement des résultats doit être rigoureux.
- *Intégrer l'épreuve dans un cadre général.* L'épreuve devrait être intégrée dans un cadre général de pratiques. Elle devrait s'aligner avec les plans d'études et les moyens d'enseignement, mais doit aussi correspondre aux pratiques des enseignants.

27.2. Clarifier/minimiser le rôle de contrôle de l'EE pour les enseignants

Par leur qualité d'objectivité et d'égalité, les évaluations externes assument, dans de nombreux contextes, le rôle de certification et/ou à d'orientation les élèves. Comme le confirme notre étude, ces fonctions créent de fortes pressions auprès des enseignants. Ces enjeux entraînent des résistances et des effets négatifs sur le terrain (*teaching to the test*, rétrécissement du curriculum). Afin d'éviter ou de minimiser ces effets, il est pour nous nécessaire de clarifier les enjeux liés aux résultats de l'EE. Différents moyens sont envisageables :

- *Limiter l'usage externe des résultats de l'épreuve.* Il est bien sûr important que les parents des élèves et les cadres de la scolarité puissent avoir un regard sur les résultats des élèves. Toutefois, les résultats de l'épreuve ne devraient pas servir à la comparaison, à des classements ou à des palmarès entre classes ou écoles. L'épreuve ne devrait pas être intégrée dans un système qui prévoit des sanctions et des récompenses. L'interprétation des résultats devrait se faire plutôt de façon critériée que normative.
- *Réduire l'importance des résultats de l'épreuve.* Les résultats de l'épreuve à eux seuls ne devraient pas servir à une prise de décision importante (certification et/ou orientation des

élèves). Ils devraient être accompagnés d'autres informations (évaluation des apprentissages par l'enseignant notamment).

- *Déterminer les responsabilités.* Les résultats de l'épreuve ne doit pas être engagée sur la seule responsabilité d'un enseignant. Cette responsabilité devrait partagée dans une chaîne de responsabilités, par exemple entre tous les enseignants de l'établissement en lien avec les moyens mis à leur disposition (révision du curriculum, soutien, supervision, etc.).
- *Faire confiance au jugement professionnel des enseignants.* Les résultats de l'évaluation interne (celle de l'enseignant) et externe (celle de l'épreuve) ne devrait pas être hiérarchisés. Davantage de confiance doit être faite aux pratiques évaluatives des enseignants, mais des moyens de perfectionnement doivent leur être fournis (collaboration, modération sociales, etc.). Des épreuves externes de référence sont d'ailleurs un moyen de développer ces compétences. Il faut également encourager les enseignants à développer leurs propres outils, ne pas les maintenir dans un état de dépendance.
- *Préciser les contenus évalués.* L'épreuve doit porter sur des objectifs clairement identifiés. En outre, ces contenus doivent toucher la globalité des contenus inscrits dans les plans d'études et des compétences globales.
- *Temporalité.* L'épreuve doit avoir lieu à la fin d'un cycle d'apprentissage, afin de laisser le temps nécessaire à l'apprentissage des contenus. Elle peut aussi avoir lieu à différents moments pour permettre une évaluation longitudinale.

27.3. Ne pas faire d'une pierre deux coups

La littérature montre que les épreuves externes remplissent bien souvent, surtout pour limiter leur nombre et leur coût, simultanément plusieurs fonctions (contrôle et/ou formation). Une seule épreuve externe suit, dans certains cas, différents objectifs (par ex. pilotage du système, orientation des élèves, contrôle des enseignants, etc.). L'intention est alors peu claire pour les acteurs du terrain ce qui crée des tensions, les finalités de l'évaluation externe étant alors contradictoires¹⁰⁴. Les chercheurs s'accordent sur le fait que les évaluations externes, pour être valides et avoir des effets positifs, ne devraient remplir qu'une seule et unique fonction. Dans le cas contraire, soit leurs objectifs ne sont soit pas remplis, soit elles entraînent des effets inattendus et indésirables (Eurydice, 2009; Mons & Crahay, 2011; Mons & Pons, 2006). Certaines études démontrent ce constat (Au, 2007; Qi, 2007; Spratt, 2005). Lorsqu'une épreuve externe suit par exemple simultanément des objectifs de

¹⁰⁴ Ces tensions sont l'objet du point 5.4 (p.45-47).

certification et/ou d'orientation des élèves et de développement des pratiques des enseignants, la première finalité prend le dessus car elle comporte davantage d'enjeux.

Dans le contexte fribourgeois, les épreuves cantonales endossent simultanément plusieurs finalités¹⁰⁵. Elles servent à la fois aux autorités, aux enseignants, aux élèves et à leurs parents. Il est intéressant d'observer spécifiquement l'épreuve d'orientation de 6P car nos analyses contrastent en partie le constat théorique précédemment formulé. En effet, cette épreuve contribue simultanément à orienter les élèves dans les filières du secondaire et à harmoniser les pratiques des enseignants. Cette épreuve tient donc une double finalité. Pourtant, nos analyses montrent que l'effet d'harmonisation des pratiques d'évaluation sommative n'est pas significativement plus faible pour ce type d'EE. Par contre, notre étude montre que l'EE d'orientation est perçue comme moins utile par les enseignants que l'EE de référence. Cette différence d'utilité est certainement due au caractère d'orientation, finalité à laquelle les enseignants donnent certainement davantage d'importance. En outre, les enseignants des classes de 5-6P n'ont pas l'occasion de bénéficier d'épreuves de référence, les administrateurs donnant ce rôle de « référence » également à l'épreuve d'orientation.

Outre les multiples finalités des épreuves fribourgeoises, leur transparence et leur communication par les autorités peuvent également être discutées. En effet, les enseignants interrogés témoignent du manque de transparence et de communication quant aux finalités des deux types d'épreuves (de référence et d'orientation). Ceci peut créer chez eux un sentiment d'incertitude ou de crainte, ne sachant pas toujours à quoi et à qui servent les résultats des élèves (orientation/information pour les élèves et leurs parents, contrôle des enseignants, pilotage du système). Aucune directive ne clarifie ces aspects. Cette situation peut donc entraîner des enjeux pour les enseignants et donc des effets indésirables sur leurs pratiques.

¹⁰⁵ Les finalités des épreuves externes fribourgeoises sont présentées au point 14.2 (p.104-110).

28. Limites et prolongements du travail

Nous concluons ce travail de thèse en relevant les limites de notre recherche. Ces dernières ouvrent des perspectives de recherche complémentaires et/ou alternatives.

Nous avons fait le choix d'une approche de recherche mixte par laquelle les enseignants ont répondu à notre questionnaire soit par entretien, soit par questionnaire. Pour ces deux étapes, les données sont issues du discours des enseignants et sont des informations auto-rapportées. Ces méthodes peuvent présenter des biais liés à la subjectivité des données. En effet, il est probable qu'il y ait un décalage entre les réponses des enseignants interrogés et leurs actions effectives en classe. Il est aussi possible que les réponses des enseignants soient marquées d'une certaine désirabilité sociale. Afin d'éviter les biais dus à nos choix méthodologiques, d'autres voies peuvent être explorées pour traiter les mêmes questions de recherche. Premièrement, il serait pertinent de faire une observation des pratiques effectives des enseignants, notamment par l'analyse de documents ou par leur observation en classe. Deuxièmement, des données auraient pu être prises de façon longitudinale en interrogeant les enseignants à plusieurs moments de l'année scolaire. Troisièmement, il aurait pu être pertinent de confronter les enseignants aux résultats de notre recherche, de manière individuelle ou sous la forme de *focus-group*.

Il serait également nécessaire de compléter les résultats de notre recherche, une partie de la variance des effets de l'évaluation externe restant inexplicée. Nous pensons tout d'abord aux facteurs liés au contexte (organisation de la scolarité, formes de politique, formation des enseignants, etc.). Il serait alors nécessaire de répliquer cette étude dans des contextes éducatifs divers (écoles de degrés différents, écoles privées, écoles gouvernées par d'autres plans d'études, systèmes scolaires incluant d'autres politiques ou dispositifs d'évaluation externe et d'*accountability*, etc.). Ceci permettrait d'ouvrir nos questions et hypothèses sur des facteurs liés au contexte (voir modèle p.88). Il est également nécessaire de compléter ou d'affiner les données sur les perceptions des enseignants quant aux épreuves externes. Certains aspects n'ayant pas directement été pris en compte (responsabilités, résistances, etc.). Enfin, notre recherche n'investigue que les effets de l'évaluation externe sur les pratiques des enseignants. Il serait utile d'observer également leurs usages et leur impact sur les élèves, leurs parents, mais aussi sur les cadres intermédiaires et les responsables politiques.

RÉFÉRENCES

- Albarello, L. (2003). *Apprendre à chercher: l'acteur social et la recherche scientifique*. Bruxelles: De Boeck.
- Alderson, J. (2004). Foreword. In Y. Watanabe & A. Curtis (Éd.), *Washback in language testing: Research contexts and methods*. London : Lawrence Erlbaum
- Alderson, J., & Wall, D. (1992). Does Washback Exist? *Applied Linguistics*, 14, 115-130.
- Allal, L. (1991). *Vers une pratique de l'évaluation formative: matériel de formation continue des enseignants*. Bruxelles: De Boeck Supérieur.
- Allal, L. (2007). Régulation des apprentissages: orientations conceptuelles pour la recherche et la pratique en éducation. In L. Allal & L. Mottier Lopez (Éd.), *Régulation des apprentissages en situation scolaire et en formation* (p. 7-23). Bruxelles: De Boeck.
- Allal, L. (2009). Pratiques évaluatives des enseignants face aux méthodologies des recherches évaluatives portant sur le système scolaire. In L. Mottier Lopez & M. Crahay, *Evaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes* (p. 29-45). Bruxelles: De Boeck.
- Allal, L., & Mottier Lopez, L. (2005). L'évaluation formative de l'apprentissage : revue de publications en langue française. In OCDE (Éd.), *L'évaluation formative. Pour un meilleur apprentissage dans les classes secondaires* (p. 241-264). Paris: OCDE.
- Amrein, A., & Berliner, D. (2002). High-Stakes Testing, Uncertainty, and Student Learning. *Education Policy Analysis Archives*, 10(18),
- Amrein-Beardsley, A., & Berliner, D. C. (2003). Re-analysis of NAEP Math and Reading Scores in States with and without High-stakes Tests: Response to Rosenshine. *education policy analysis archives*, 11(25).
- Au, W. (2007). High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. *Educational Researcher*, 36(5), 258-267.
- Baeriswyl, F., & Périsset Bagnoud, D. (2008). *Bildungsstandards kontrovers. Les standards de formation : controverses*. Fribourg: Academic Press.
- Baeriswyl, F., Trautwein, U., Wandeler, C., & Lüdtke, O. (2009). Wie gut prognostizieren subjective Lehrerempfehlungen und schulische Testleistungen beim Übertritt die Mathematik- und Deutschleistung in der Sekundarstufe I? *Zeitschrift für Erziehungswissenschaft*, 12, 352-372.
- Baeriswyl, F., Wandeler, C., & Trautwein, U. (2011). « Auf einer anderen Schule oder bei anderen Lehrkraft hätte es für's Gymnasium gereicht »: Eine Untersuchung zur Bedeutung von Schulen und Lehrkräften für die Übertrittsempfehlung. *Zeitschrift für Pädagogische Psychologie*, 25(1), 39-47.
- Baeriswyl, F., Wandeler, C., Trautwein, U., & Oswald, K. (2006). Leistungstest, Offenheit von Bildungsgängen und obligatorische Beratung von Eltern. *Zeitschrift für Erziehungswissenschaft*, 3, 374-392.
- Bandura, A. (2003). *Auto-efficacité: Le sentiment d'efficacité personnelle*. Bruxelles: De Boeck Université.

- Behrens, M. (2005). Analyse de la littérature critique sur le développement, l'usage et l'implémentation de standards dans un système éducatif. Neuchâtel: IRDP.
- Behrens, M. (2010). Standards ou les escaliers mouvants de Poudlard. In P. Gilliéron Giroud & L. Ntamakiliro (Éd.), *Réformer l'évaluation scolaire : mission impossible ?* (p. 113-129). Berne: Peter Lang.
- Beillerot, J. (2005). L'obligation de résultats en éducation : une méthode de formation de l'opinion publique. In C. Lessard & P. Meirieu (Éd.), *L'obligation de résultats en éducation : évolutions, perspectives et enjeux internationaux* (p. 71-78). Bruxelles: De Boeck Université.
- Bélaïr, L. (2005). Les dérives de l'obligation de résultats ou l'art de surfer sans planche. In C. Lessard & P. Meirieu (Éd.), *L'obligation de résultats en éducation : évolutions, perspectives et enjeux internationaux* (p. 180-187). Bruxelles: De Boeck Université.
- Berliner, D. C., & Nichols, S. (2007). High-stakes testing is putting the nation at risk. *Education Week*, 26(27), 36-48.
- Bishop, J. H. (1995). The impact of Curriculum-Based External Examinations on School Priorities and Student Learning. *International Journal of Educational Research*, 23(8), 653-752.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, 21(1), 5-31. <http://doi.org/10.1007/s11092-008-9068-5>
- Blais, J.-G. (2005). L'obligation de résultats à la lumière de l'interaction entre le quantitatif et le social. In C. Lessard & P. Meirieu (Éd.), *L'obligation de résultats en éducation : évolutions, perspectives et enjeux internationaux* (p. 124-144). Bruxelles: De Boeck Université.
- Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). *Evaluation to improve learning*. McGraw-Hill.
- Booher-Jennings, J. (2005). Below the Bubble: « Educational Triage » and the Texas Accountability System. *American Educational Research Journal*, 42(2), 231-268.
- Bottani, N. (2004). Les évaluations internationales des acquis des élèves et leur impact sur les politiques de l'éducation. *Evaluation des résultats des élèves: quels impacts sur les politiques d'éducation?*, 11(2), 9-20.
- Bransford, J., Brown, A. L., & Cocking, R. R. (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington: National Academy Press.
- Broadfoot, P. (2000). Un nouveau mode de régulation dans un système décentralisé : l'Etat évaluateur. *Revue française de pédagogie*, 130(1), 43-55.
- Buisson-Fenet, H., & Pons, X. (2011). *Les pratiques d'évaluation externe des établissements scolaires en France, au Royaume-Uni et en Suisse : vers des figures de l'État éducateur contemporain en Europe*.
- Carette, V. (2007). Les implications de la notion de compétence sur l'évaluation. *Education - Formation*, e-286, 51-61.

- Carette, V. (2009). Et si on évaluait des compétences en classe ? À la recherche du « cadrage instruit ». In L. Mottier Lopez & M. Crahay (Éd.), *Evaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. (p. 147-163). Bruxelles: De Boeck.
- Carette, V., & Dupriez, V. (2009). La lente émergence d'une politique scolaire en matière d'évaluation des élèves: Quinze ans de transformations en Belgique francophone. *Mesure et évaluation en éducation*, 32(3), 25-46.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305-331.
- Carricano, M., Bertrandias, L., & Poujol, F. (2010). *Analyse de données avec SPSS®*. Paris: Pearson Education.
- Cattonar, B., Dumay, X., Mangez, C., & Maroy, C. (2010). *Evaluations externes dans l'enseignement primaire en Belgique francophone : Réceptions et usages d'outils de régulation basés sur les connaissances*. Bruxelles: European Commission within the Sixth Framework Program.
- CDIP (2004). *HARMOS : Finalités et conception du projet* (CDIP.). Berne: CDIP.
- Chapman, D. W., & Snyder, C. W. (2000). Can high stakes national testing improve instruction: reexamining conventional wisdom. *International Journal of Educational Development*, 20(6), 457-474.
- Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and teacher education*, 15(3), 253-272.
- Cizek, G. (2001). More Unintended Consequences of High-Stakes Testing. *Educational Measurement: Issues and Practice*, 20(4), 19-27.
- Coburn, C. E. (2004). Beyond Decoupling: Rethinking the Relationship Between the Institutional Environment and the Classroom. *Sociology of Education*, 77(3), 211-244.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. London: Erlbaum Associates.
- Crahay, M. (2009). Articuler l'évaluation en classe et le pilotage des systèmes, est-ce possible ? In L. Mottier Lopez & M. Crahay (Éd.), *Evaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. (p. 233-251). Bruxelles: De Boeck.
- Crahay, M., & Lafontaine, D. (2012). Pistes pour une école juste et efficace. In M. Crahay (Dir.), *L'école peut-elle être juste et efficace?* (p. 455-477). Bruxelles: De Boeck.
- Crahay, M., Wanlin, P., Issaeiva, E., & Laduron, I. (2011). Fonctions, structuration et évolution des croyances (et connaissances) des enseignants. *Revue française de pédagogie*, 172(3), 85-129.
- Crocker, L. M., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason: Cengage Learning.
- Darling-Hammond, L. (1994). Performance-Based Assessment and Educational Equity. *Harvard Educational Review*, 64(1), 5-30.

- Darling-Hammond, L., & Wise, A. E. (1985). Beyond Standardization: State Standards and School Improvement. *The Elementary School Journal*, 85(3), 315-336.
- Davaud, C., & Cardinet, J. (1992). La problématique des épreuves communes. In D. Laveault (Dir.), *Les pratiques évaluatives en éducation* (p. 169-181). Montréal: Editions de l'ADMEE.
- Debard, R., & Kubow, P. K. (2002). From Compliance to Commitment: The Need for Constituent Discourse in Implementing Testing Policy. *Educational Policy*, 16(3), 387-405.
- De Landsheere, G. D. (1979). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris: Presses Universitaires de France.
- Demailly, L. (2001). *Évaluer les politiques éducatives: sens, enjeux, pratiques*. Bruxelles: De Boeck Supérieur.
- Dierendonck, C. (2008). *Comment les évaluations externes des acquis des élèves sont-elles perçues par les enseignants du primaire dans les cantons de Neuchâtel, Vaud et Fribourg? Enquête exploratoire*. Neuchâtel: IRDP.
- Döbert, H., & Dederling, K. (2008). *Externe Evaluation von Schulen: historische, rechtliche und vergleichende Aspekte*. Münster: Waxmann Verlag.
- Dumay, X., Cattonar, B., Maroy, C., & Mangez, C. (2013). Modalités de mise en oeuvre des instruments d'évaluation et institutionnalisation de la régulation par les résultats dans les établissements scolaires. In Maroy C. (Dir), *L'école à l'épreuve de la performance. Les politiques de régulation par les résultats* (p. 125-139). Bruxelles: De Boeck.
- Dupriez, V. (2004). La place de l'évaluation comme ressource pour le pilotage des systèmes scolaires: état des lieux en Belgique francophone et en Angleterre. *Les cahiers de Recherche en Education et Formation*, 35, 1-18.
- Dupriez, V. (2005). Régulation et pilotage des systèmes scolaires: Les voies de l'harmonisation dans deux systèmes scolaires historiquement décentralisés, l'Angleterre et la Belgique francophone. *Schweizerische Zeitschrift für Bildungswissenschaften*, 27(1), 37-52.
- Duran, P., & Monnier, É. (1992). Le développement de l'évaluation en France. Nécessités techniques et exigences politiques. *Revue française de science politique*, 42(2), 235-262.
- Duru-Bellat, M., & Jarousse, J.-P. (2001). Portée et limites d'une évaluation des politiques et des pratiques éducatives par les résultats. *Education et sociétés*, 8(2), 97-109.
- Dussault, M., Villeneuve, P., & Deaudelin, C. (2001). L'échelle d'autoefficacité des enseignants: validation canadienne-française du Teacher efficacy scale. *Revue des sciences de l'éducation*, 27(1), 181-194.
- Eurydice. (2009). *Les évaluations standardisées des élèves en Europe: objectifs, organisation et utilisation des résultats*. Bruxelles: Agence exécutive «Éducation, audiovisuel et culture» (EACEA P9 Eurydice).
- Felouzis, G., & Hanhart, S. (2011). Politiques éducatives et évaluation: nouvelles tendances, nouveaux acteurs. In G. Felouzis & S. Hanhart (Éd.), *Gouverner l'éducation par les nombres ? Usage, débats et controverses* (p.7-31). Bruxelles: De Boeck.

- Fuller, E. J., & Treisman, P. U. (2001). Comment by Philip Uri Treisman and Edward J. Fuller. *Brookings Papers on Education Policy*, 20(1), 208-218.
- Grisay, A. (1984). Les mirages de l'évaluation scolaire (1). *Revue de la Direction Générale de l'Organisation des Etudes*, 17(5), 29-42.
- Grissmer, D., & Flanagan, A. (1998). Exploring Rapid Achievement Gains in North Carolina and Texas. Lessons from the States. Washington : National education goals panel.
- Hadji, C. (1989). *L'Évaluation, règles du jeu: des intentions aux outils*. Paris: ESF.
- Haney, W. (2000). The Myth of the Texas Miracle in Education. *Education Policy Analysis Archives*, 8(41).
- Haney, W., & Madaus, G. (1986). Effects of Standardized Testing and The Future of The National Assessment of Educational Progress. Working paper for the NAEP. Chestnut Hill: Study Group on the NAEP.
- Hanushek, E. A., & Raymond, M. E. (2004). Does School Accountability Lead to Improved Student Performance? *Journal of Policy Analysis*, 24(2), 297-327.
- Harris, D. N., & Herrington, C. D. (2006). Accountability, Standards, and the Growing Achievement Gap: Lessons from the Past Half-Century. *American Journal of Education*, 112(2), 209-238.
- Helgoy, I., & Homme, A. (2007). Towards a New Professionalism in School? A Comparative Study of Teacher Autonomy in Norway and Sweden. *European Educational Research Journal*, 6(3), 232-249.
- Hong, W.-P., & Youngs, P. (2008). Does High-stakes Testing increase Cultural Capital among Low-income and racial Minority Students? *Education Policy Analysis Archives*, 16(6), 1-36.
- Huguenin, J.-M., & Soguel, N. (2013). Les systèmes d'indicateurs des établissements scolaires de Suisse : quelle cohérence avec les déterminants de la performance scolaire ? In C. Maroy (Dir.), *L'école à l'épreuve de la performance. Les politiques de régulation par les résultats* (p. 157-180). Bruxelles: De Boeck Supérieur.
- Issaieva, E., & Crahay, M. (2010b). *Conceptions de l'évaluation, de l'apprentissage, de l'intelligence et sentiments d'auto-efficacité chez les enseignants du primaire : Quels liens ?* Actes du congrès de l'AREF, septembre 2010, Université de Genève.
- Issaieva, E., & Crahay, M. (2010b). Conceptions de l'évaluation scolaire des élèves et des enseignants: validation d'échelles et étude de leurs relations. *Mesure et évaluation en éducation*, 33(1), 31-62.
- Issaieva, É., Pini, G., & Crahay, M. (2011). Positionnements des enseignants et des élèves du primaire face à l'évaluation : une convergence existe-t-elle ? *Revue française de pédagogie*, 176, 5-26.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33(7), 14-26.
- Jones, B. (2007). The Unintended Outcomes of High-Stakes Testing. *Journal of Applied School Psychology*, 23(2), 65-86.

- Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *The Phi Delta Kappan*, 81(3), 199–203.
- Jones, M. G., Jones, B., & Hargrove, T. (2003). *The unintended consequences of high-stakes testing*. Oxford: Rowman & Littlefield
- Klieme, E., Avenarius, H., Blum, W., Döbrich, W., Gruber, H., & Prenzel, M. (2004). *Le développement de standards nationaux de formation. Une expertise*. Bonn: Ministère fédéral de l'Éducation et de la Recherche.
- Koretz, D. (2005). Alignment, High Stakes, and the Inflation of Test Scores. *Yearbook of the National Society for the Study of Education*, 104(2), 99–118.
- Lafontaine, D. (2000). Les évaluations externes en communauté française de Belgique: Quels sens donner à la mesure? *Cahiers du Service de Pédagogie expérimentale - Université de Liège*, 1-2, 110-114.
- Lafontaine, D. (2012). *Des politiques aux pratiques d'évaluation en Belgique francophone: la cohérence en questions*. Actes du 20ème colloque de l'ADMEE-Europe, janvier 2012, Luxembourg.
- Lafontaine, D., Soussi, A., & Nidegger, C. (2009). Evaluations internationales et/ou épreuves nationales : tensions et changement de pratiques. In Mottier Lopez L. & Crahay M. (Dir.), *Evaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes* (p. 61-80). Bruxelles: De Boeck.
- Landry, F. (2005). *Les conditions de promotion et d'orientation au cours de la scolarité obligatoire en Suisse romande et au Tessin: pratiques cantonales concernant l'organisation d'épreuves, d'examens, de tests ainsi que l'obtention de certificats ou diplômes : année scolaire 2005-2006*. Neuchâtel: IRDP.
- Landsheere, G. D. (1994). *Le pilotage des systèmes d'éducation*. Bruxelles: De Boeck Université.
- Laveault, D. (2007). De la « régulation » au « réglage »: élaboration d'un modèle d'autoévaluation des apprentissages. In L. Allal & L. Mottier Lopez (Éd.), *Régulation des apprentissages en situation scolaire et en formation* (p. 207-234). Bruxelles: De Boeck.
- Laveault, D. (2009a). L'amélioration de l'efficacité du système éducatif: sur quels indicateurs s'appuyer? In X. Dumay & V. Dupriez (Éd.), *L'efficacité dans l'enseignement. Promesses et zones d'ombre*. (p. 177-196). Bruxelles: De Boeck.
- Laveault, D. (2009b). L'évaluation en classe: des politiques aux pratiques. *Mesure et évaluation en éducation*, 32(3), 3-22.
- Laveault, D. (2012a). *L'évaluation des compétences en classe : des politiques aux pratiques*. Actes du 24ème colloque de l'ADMEE-Europe, janvier 2012, Luxembourg.
- Laveault, D. (2012b). Soixante ans de bons et mauvais usages du alpha de Cronbach. *Mesure et évaluation en éducation*, 35(2), 1-7.
- L'Écuyer, R. (1990). *Méthodologie de l'analyse développementale de contenu: méthode GPS et concept de Soi*. Québec: PUQ.

- Lee, J. (2008). Is Test-Driven External Accountability Effective? Synthesizing the Evidence From Cross-State Causal-Comparative and Correlational Studies. *Review of Educational Research*, 78(3), 608-644.
- Lessard, C., & Meirieu, P. (2005). *L'obligation de résultats en éducation: Évolutions, perspectives et enjeux internationaux*. Bruxelles: De Boeck Université.
- Linn, R. L. (2000). Assessments and Accountability. *Educational Researcher*, 29(2), 4-16.
- Longchamp, A.-L. (2010). *Epreuves cantonales de référence de mathématiques en 6e année. Point de vue d'enseignants sur l'utilisation du dispositif*. Lausanne: URSP.
- Looney, J. (2009). *Assessment and Innovation in Education*. OECD Education Working Papers.
- Madaus, G. F. (1991). The effects of important tests on students: Implications for a national examination system. *The Phi Delta Kappan*, 73(3), 226-231.
- Madaus, G. F., & Kellaghan, T. (1992). Curriculum Evaluation and Assessment. In P. W. Jackson (Dir.), *Handbook of Research on Curriculum* (p. 119-154). New York: MacMillan.
- Malet, R. (2011). Un analyseur des formes de redéfinition stratégiques de l'Etat: les politiques d'imputabilité en direction des enseignants en Angleterre et aux Etats-Unis. In V. Dupriez & N. Mons (Dir.), *Les politiques d'accountability. Du changement institutionnel aux transformations locales* (p. 35-60). Association francophone d'Education comparée.
- Marc, V., & Wirthner, M. (2012). *Épreuves romandes communes : de l'analyse des épreuves cantonales à un modèle d'évaluation adapté au PER*. Rapport final du projet EpRoCom. Neuchâtel: IRDP.
- Maroy, C. (2004). Régulation et évaluation des résultats des systèmes d'enseignement. *Politiques d'Education et de Formation*, 11(2), 21-36.
- Maroy, C. (2005a). Les évolutions du travail enseignant en Europe. Facteurs de changement, incidences et résistances. *Les Cahiers de Recherche en Education et changement*, 42, 1-35.
- Maroy, C. (2005b). Vers une régulation post-bureaucratique des systèmes d'enseignement en Europe? *Les cahiers de Recherche en Education et Formation*, 49, 1-30.
- Maroy, C. (2009). Régulation post-bureaucratique des systèmes d'enseignement et travail enseignant. In L. Mottier Lopez & M. Crahay (Dir.), *Evaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. (p. 83-100). Bruxelles: De Boeck.
- Maroy, C. (2012a). Les politiques d'accountability au service de la confiance dans l'institution scolaire et les enseignants? *Revue suisse des sciences de l'éducation*, 34(1), 57-70.
- Maroy, C. (2012b). Performance scolaire et régulation par les connaissances en éducation. *Recherches sociologiques et anthropologiques*, (43-2), 1-13.
- Maroy, C. (2013). Politiques et outils de «l'école de la performance»: accountability, régulation par les résultats et pilotage. In C. Maroy (Éd.), *L'école à l'épreuve de la*

- performance. Les politiques de régulation par les résultats* (p. 14-31). Bruxelles: De Boeck.
- Maroy, C. (2014). La confiance des sous contrôle: le professionnalisme enseignant face aux nouvelles régulations. In *Enseigner: un métier sous contrôle? Entre autonomie professionnelle et normalisation du travail* (p. 209-214). Issy-les-Moulineaux: ESF.
- Maroy, C., & Cattonar, B. (2002). Professionnalisation ou déprofessionnalisation des enseignants? Le cas de la Communauté française de Belgique. *Les cahiers de Recherche en Education et Formation*, 18, 1-29.
- Maroy, C., & Dupriez, V. (2000). La régulation dans les systèmes scolaires. *Revue française de pédagogie*, 130(1), 73-87.
- Maroy, C., & Mangez, C. (2011). La construction des politiques d'évaluation et de pilotage du système scolaire en Belgique francophone: nouveau paradigme politique et médiation des experts. In G. Felouzis & S. Hanhart (Dir.), *Gouverner l'éducation par les nombres ?* (p. 53-76). Bruxelles: De Boeck.
- McDonnell, L. M. (1994). Assessment Policy as Persuasion and Regulation. *American Journal of Education*, 102(4), 394-420.
- McMillan, J. H. (2005). The Impact of High-Stakes Test Results on Teachers. *Online Submission*, 27, 1-26.
- Meirieu, P. (2005). L'école entre la pression consumériste et l'irresponsabilité sociale. In C. Lessard & P. Meirieu (Éd.), *L'obligation de résultats en éducation: évolutions, perspectives et enjeux internationaux* (p. 5-21). Bruxelles: De Boeck Université.
- Meuret, D. (2012). *Les effets de l'accountability sur les politiques d'éducation aux Etats-Unis*. Dijon: IREDU
- Miles, M., & Huberman, A. M. (2010). *Analyse des données qualitatives*. Bruxelles: De Boeck Université.
- Mizutani, S. (2009). *The mechanism of washback on teaching and learning*. Thèse de doctorat, University of Auckland, Auckland.
- Monseur, C., & Demeuse, M. (2004). Quelques réflexions méthodologiques à propos des enquêtes internationales dans le domaine de l'éducation. *Politiques d'Éducation et de formation*, 11, 37-54.
- Monseur, C., & Demeuse, M. (2005). Les évaluations externes permettent-elles une régulation efficace. In A. Baye, M. Demeuse, J. Nicaise & M.-H. Straeten (Eds.), *Vers une école juste et efficace* (p. 489-517). Bruxelles: De Boeck.
- Mons, N. (2007b). *Les nouvelles politiques éducatives*. Paris: PUF.
- Mons, N. (2007c). *Les nouvelles politiques éducatives: La France fait-elle les bons choix ?* Paris: PUF.
- Mons, N. (2007a). L'évaluation des politiques éducatives. Apports, limites et nécessaire renouvellement des enquêtes internationales sur les acquis des élèves. *Revue internationale de politique comparée*, 14(3), 409-423.
- Mons, N. (2008a). Autonomie et contrôle du travail enseignant. *Revue internationale d'éducation de Sèvres*, 48, 109-117.

- Mons, N. (2008b). Évaluation des politiques éducatives et comparaisons internationales. *Revue française de pédagogie. Recherches en éducation*, 164, 5-13.
- Mons, N. (2009). Effets théoriques et réels des politiques d'évaluation standardisée. *Revue française de pédagogie. Recherches en éducation*, 169, 99-140.
- Mons, N. (2013). Evaluation standardisée des élèves et inégalités scolaires d'origine sociale: discours théoriques et réalité empirique. In C. Maroy (Dir.), *L'école à l'épreuve de la performance. Les politiques de régulation par les résultats* (p. 33-49). Bruxelles: De Boeck.
- Mons, N., & Crahay, M. (2011). L'évaluation des performances scolaires des élèves : un instrument d'évaluation des politiques éducatives ? In G. Felouzis & S. Hanhart (Eds.), *Gouverner l'éducation par les nombres ? Usage, débats et controverses*. (p. 77-98). Bruxelles: De Boeck.
- Mons, N., & Dupriez, V. (2010). Les politiques d'accountability. *Recherche & formation*, 65(3), 45-59.
- Mons, N., & Durpiez, V. (2010). Les politiques d'accountability. *Recherche et formation*, (65), 45-59.
- Mons, N., & Pons, X. (2006). *Les standards en éducation dans le monde francophone: Une analyse comparative*. Neuchâtel: IRDP.
- Mottier Lopez, L. (2009). L'évaluation en éducation : des tensions aux controverses. In L. Mottier Lopez & M. Crahay (Dir.), *Evaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. (p. 7-25). Bruxelles: De Boeck.
- 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.
- Mucchielli, R. (2006). *L'analyse de contenu des documents et des communications*. Paris: ESF.
- Newton, P. E. (2007). *Evaluating assessment systems*. Carrickfergus: Qualifications and Curriculum Authority
- Nichols, S. L. (2007). High-Stakes Testing: Does It Increase Achievement? *Journal of Applied School Psychology*, 23(2), 47-64.
- Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-Stakes Testing and Student Achievement: Does Accountability Pressure Increase Student Learning? *Education Policy Analysis Archives*, 14.
- Normand, R. (2005). La mesure de l'école : politique des standards et management par la qualité. *Cahiers de la recherche sur l'éducation et les savoirs*, 1, 67-82.
- Normand, R., & Derouet, J.-L. (2011). Évaluation, développement professionnel et organisation scolaire. *Revue française de pédagogie*, 174(1), 5-20.
- Ntamakiliro, L. (2003). La place des épreuves communes standardisées dans le système d'évaluation des élèves. *Deux points ouvrez les guillemets*, 19, 19-23.

- Ntamakiliro, L., & Longchamp, A.-L. (2008). *Aspects de la standardisation de l'évaluation externe dans le canton de Vaud*. Actes du 20ème colloque de l'ADMEE, janvier 2008, Genève.
- Ntamakiliro, L., & Moreau, J. (2011). *Analyse des résultats des élèves vaudois aux épreuves cantonales de référence*. Lausanne: URSP.
- Ntamakiliro, L., & Tessaro, W. (2002). *La perception de l'évaluation externe des élèves par les enseignants primaires genevois*. Actes du 15ème colloque de l'ADMEE et congrès annuel de la SSRE, janvier 2002, Lausanne.
- Ntamakiliro, L., & Tessaro, W. (2010). Plus d'épreuves externes : vers une standardisation de l'évaluation scolaire ? In P. Gilliéron Giroud & L. Ntamakiliro (Éd.), *Réformer l'évaluation scolaire: mission impossible ?* (p. 165-185). Berne: Peter Lang.
- Ntamakiliro, L., & Tessaro, W. (2010b). Vers plus d'épreuves externes: une standardisation de l'évaluation scolaire ? In *Réformer l'évaluation scolaire: mission impossible ?* (P. Gilliéron & L. Ntamakiliro, p. 165-185). Berne: Peter Lang.
- OCDE (2007). *PISA 2006: Science Competencies for Tomorrow's World: Volume 1: Analysis*. OECD Publishing.
- Osborn, M. (2006). Changing the context of teachers' work and professional development: A European perspective. *International Journal of Educational Research*, 45(4-5), 242-253.
- Pajares, M. F. (1992). Teachers' Beliefs and Educational Research: Cleaning up a Messy Construct. *Review of Educational Research*, 62(3), 307-332.
- Pan, Y.-C. (2009). A review of washback and its pedagogical implications. *VNU Journal of Science, Foreign Languages*, 25, 257-263.
- Paul, J.-J. (2005). L'économie de l'éducation permet-elle de légitimer l'obligation de résultats ? In C. Lessard & P. Meirieu (Eds.), *L'obligation de résultats en éducation : évolutions, perspectives et enjeux internationaux* (p. 79-87). Bruxelles: De Boeck Université.
- Petrucci, F. (2010). Compétences en lecture à Genève: influence des caractéristiques des élèves et des classes à différents moments de la scolarité obligatoire. *Mesure et évaluation en éducation*, 33(1), 79-106.
- Pluye, P., Nadeau, L., Ggnon, M.-P., Grad, R., Johnson-Lafleur, J., & Griffiths, F. (2012). Les méthodes mixtes. In C. Dagenais & V. Ridde (Eds.), *Approches et pratiques en évaluation de programme* (p. 123-143). Montréal: Les Presses de l'Université de Montréal.
- Pons, X. (2011). *L'évaluation des politiques éducatives*. Paris: PUF.
- Popham, W. J. (1987). The Merits of Measurement-Driven Instruction. *The Phi Delta Kappan*, 68(9), 679-682.
- Popham, W. J. (2001). Teaching to the Test? *Educational Leadership*, 58(6), 16-21.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142-173.

- Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education: Principles, Policy & Practice*, 14(1), 51-74.
- Ramirez, J. (2011). L'évaluation des élèves; entre externalisation et hétérogénéité. In G. Felouzis & S. Hanhart (Eds.), *Gouverner l'éducation par les nombres ? Usage, débats et controverses* (p. 185-200). Bruxelles: De Boeck.
- Rey, B. (2005). La notion de compétence permet-elle de répondre à l'obligation de résultats dans l'enseignement ? In C. Lessard & P. Meirieu (Dir.), *L'obligation de résultats en éducation : évolutions, perspectives et enjeux internationaux* (p. 233-242). Bruxelles: De Boeck.
- Roegiers, X. (2004). *L'école et l'évaluation: des situations pour évaluer les compétences des élèves*. Bruxelles: De Boeck Supérieur.
- Rosenshine, B. (2003). High-Stakes Testing: Another Analysis. *Education Policy Analysis Archives*, 11(24), 1-34.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144. <http://doi.org/10.1007/BF00117714>
- Scallon, G. (1988). *L'évaluation formative des apprentissages*. Laval: Presses Université.
- Scallon, G. (1991). *L'évaluation formative des apprentissages*. Laval: Presses Université.
- Shepard, L. (2003). The Hazards of High-Stakes Testing. *Issues in Science and Technology*, 19(2), 53-58.
- Shih, C. M. (2009). How tests change teaching: A model for reference. *English Teaching: Practice and Critique*, 8(2), 188-206.
- Shohamy, E. (2001). The Power of Tests: A Critical Perspective on the Uses of Language Tests. *Journal of Writing Assessment*, 3(1), 55-60.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: washback effect over time. *Language Testing*, 13(3), 298-317.
- Smith, L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11.
- Soussi, A., Ducrey, F., Ferrez, E., Nidegger, C., & Viry, G. (2006). *EVALECOPO. Pratiques d'évaluation: ce qu'en disent les enseignants (à l'école obligatoire et dans l'enseignement postobligatoire général)*. Genève: Département de l'instruction publique - Service de la recherche en éducation.
- Soussi, A., & Nidegger, C. (2012). *Les évaluations externes dans quelques pays ou régions d'Europe : élaboration, analyses et diffusion* Note d'information du SRED No. 54. Genève: Département de l'instruction publique - Service de la recherche en éducation.
- Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
- Suchaut, B. (2008). Evaluation externe et recherche empirique en éducation: Quels apports pour l'amélioration de l'école primaire? *Recherches & éducations*, 1, 89-104.

- Tessaro, W., & Ntamakiliro, L. (2010a). *Comment les enseignants s'accommodent-ils de l'évaluation externe?* Actes du congrès de l'Actualité de la recherche en éducation et en formation (AREF), 2010, Genève.
- Tessaro, W., & Ntamakiliro, L. (2010b). *Les évaluations externes en Suisse romande: enjeux, pratiques et effets.* Actes du congrès de la section suisse de l'ADMEE-Europe, 2010, Martigny.
- Thélot, C. (2002). Evaluer l'Ecole. *Études*, 397(10), 323-334.
- Trudel, L., Simard, C., & Vonarx, N. (2007). La recherche qualitative est-elle nécessairement exploratoire? *Recherches Qualitatives*, 5, 38-45.
- Tschannen-Moran, M., & Woolfolk Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and teacher education*, 17(7), 783-805.
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning: can this be predicted or controlled? *System*, 28(4), 499-509.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Éds.), *Washback in language testing* (p. 19-36). New Jersey: Lawrence Erlbaum.
- Weiss, J. (2002). *L'évaluation externe dans un concept général et cohérent d'évaluation.* Actes du 15e colloque international de l'ADMEE-Europe et Congrès annuel de la SSRE, 2002, Lausanne.
- Wirthner, M. (2010). Un bilan des réformes de l'évaluation du travail des élèves en Suisse romande. In P. Gilliéron Giroud & L. Ntamakiliro (Éds.), *Réformer l'évaluation scolaire: mission impossible ?* (p. 13-31). Berne: Peter Lang.
- Wirthner, M., & Ntamakiliro, L. (2008). *Des épreuves de références au service de l'évaluation des enseignants. Quelle évaluation des enseignants au service de l'école?* Actes du séminaire AIDEP, 6 et 7 décembre 2007, Leysin.
- Wößmann, L. (2007). International Evidence on School Competition, Autonomy, and Accountability: A Review. *Peabody Journal of Education*, 82(2-3), 473-497.

ANNEXES

Annexe A : Canevas d'entretien.....	248
Annexe B : Arborescence finale des catégories d'analyse	252
Annexe C : Lettre aux Responsables d'établissement.....	256
Annexe D : Questionnaire	257
Annexe E : Corrélations entre les croyances des enseignants à propos de l'enseignement	264
Annexe F : Curriculum vitae	265
Annexe G : Déclaration sur l'honneur.....	266

Canevas d'entretien

1. Profil général de l'enseignant

- 1.1. Quel âge avez-vous ?
- 1.2. Depuis combien d'années (quelle année) êtes-vous enseignant du degré primaire ?
- 1.3. Dans quelles écoles avez-vous enseigné ?
- 1.4. Quel a été votre parcours dans les degrés ?
- 1.5. En quel degré enseignez-vous cette année ?
- 1.6. Depuis combien de temps enseignez-vous dans ce degré ?
- 1.7. Comment est votre classe de cette année ? Type de classe actuellement ?
(nombre, niveau, intégrations...)

2. Conceptions et pratiques de l'évaluation scolaire

- 2.1. Qu'est-ce que l'évaluation pour vous ?
- 2.2. A quoi sert-elle ?
- 2.3. Comment organisez-vous l'évaluation lors d'une séquence (thème) ?
- 2.4. Définissez-vous votre pratique d'évaluation plutôt comme globale (sur plusieurs thèmes, de grands contenus...) ou plutôt par petits pas, plus ciblés ? Pourquoi ce choix ?
- 2.5. Quand et comment pratiquez-vous l'évaluation *formative* ?
- 2.6. Quand et comment pratiquez-vous l'évaluation *sommative* ?
- 2.7. Comment construisez-vous vos évaluations (par objectifs, choix des exercices, gestion du temps...) ?
- 2.8. Quels types de questions posez-vous (vrai ou faux, questions ouvertes, recherche dans des documents...) ? (connaissances VS compétences, niveaux taxonomiques) (différences entre les branches)
- 2.9. Comment, sur quels documents vous référez vous pour définir les cibles d'évaluation ? (PER, méthodologie...)

3. Expérience(s) et perceptions des évaluations externes

- 3.1. Quelles expériences d'évaluations externes avez-vous ? (TAC, PPO, épreuves communes, PISA...)
- 3.2. Que vous évoquent ces évaluations externes ?
- 3.3. Quelles sont les buts de ces évaluations selon vous, pour : A quoi servent-elles pour :
 - pour les élèves ?
 - les enseignants ?
 - l'Etat (DICS) ?
- 3.4. Quels sont les avantages ?
 - pour les élèves ?
 - les enseignants ?
 - l'Etat (DICS) ?
- 3.5. et les inconvénients ? idem

4. Effets de l'évaluation externe sur la gestion du curriculum

- 4.1. Comment organisez-vous l'année scolaire (les contenus) lorsque vous êtes confrontés à une évaluation externe ? (Comme les autres années ?) (déjà en 5P ?) (Comme les autres années ?)
- 4.2. Que changent ces évaluations externes sur les programmes d'une année, pour vous ?
- 4.3. L'évaluation externe vous empêche-t-elle de traiter certains thèmes, certaines branches ? Autres activités ?
- 4.4. Estimez-vous avoir le temps de voir les thèmes en profondeur ?
 - Ceux de l'épreuve
 - Les autres
- 4.5. Les évaluations correspondent-elles au programme de l'année ? (et des autres années ?)
- 4.6. Que pensez-vous des thèmes abordés ? (choix, quantité, diversité...)
- 4.7. Faudrait-il évaluer toutes les branches ?

5. Effets de l'évaluation externe sur les pratiques d'enseignement-apprentissage

- 5.1. De quelle manière travaillez/organisez-vous les thèmes des évaluations externes ?
- 5.2. Vous adaptez-vous, adaptez-vous votre enseignement, d'après ces épreuves ?

- 5.3. Traitez-vous les thèmes de l'évaluation externe comme les autres ? (plus de drill, en projet, interdisciplinarité...)
- 5.4. Ces évaluations vous poussent-elles à changer vos pratiques ?
- 5.5. Estimez-vous enseigner de la même manière si vous êtes dans une année d'évaluation externe ?
- 5.6. Ces évaluations vous poussent-elles à innover, à inventer de nouvelles choses ?
- 5.7. Estimez-vous que le matériel didactique officiel soit suffisant pour préparer ces évaluations ?
Devez-vous créer du nouveau matériel ?
- 5.8. Ces évaluations vous poussent-elles à collaborer avec des collègues, des groupes de travail, à vous ouvrir ?
- 5.9. Durant une année d'évaluation externe est-il possible de faire autant d'évaluation formative ? De la différenciation ?
- 5.10. Le rythme d'apprentissage des élèves peut-il être respecté lors d'une de ces années ?

6. Effets de l'évaluation externes sur les pratiques d'évaluation interne ?

- 6.1. Comment évaluez-vous les thèmes qui sont seront traités dans l'évaluation externe ?
- 6.2. Différemment des autres thèmes ?
- 6.3. Votre pratique d'évaluation change-t-elle lors des années d'évaluation externe ?
- 6.4. Quelles sont les différences entre votre manière d'évaluer en classe (interne) et les évaluations externes (compétences, connaissances...)
- 6.5. Essayez-vous de prendre les mêmes méthodes/conditions d'évaluation que celles des épreuves externes ?

(Gestion du temps, épreuves longues, coupée en plusieurs fois ?)
- 6.6. Les évaluations externes vous donnent-elles de nouvelles idées pour votre pratique d'évaluation ?
- 6.7. Avec le temps, votre expérience des évaluations externes a-t-elle changé votre approche d'évaluation interne ?
- 6.8. Trouvez-vous que l'évaluation externe complète votre évaluation interne (différenciation) ? Va-t-elle à l'encontre ? Y a-t-il un lien entre les deux ?
- 6.9. Laquelle des deux évaluations, interne ou externe, donne la meilleure vision des aptitudes de l'élèves ?

- 6.10. Évaluent-elles des choses différentes ?
- 6.11. Pour vous, quelle est l'importance hiérarchique entre les deux ?
- 6.12. Qui sont les élèves qui réussissent le mieux les évaluations externes ? Les mêmes qu'aux évaluations internes ?

7. Pratiques de préparation aux évaluations externes

- 7.1. Comment préparez-vous vos élèves à passer des épreuves externes ?
- 7.2. Pratiquez-vous des exercices spéciaux ? (épreuves anciennes)
- 7.3. Préparez-vous les élèves à répondre de certaines façons, un travail réflexif ?
- 7.4. Quel devrait être le rôle du maître, son attitude ?
- 7.5. Quel est la part du maître dans la réussite de ses élèves ?
- 7.6. Comment vous préparez-vous vous-même aux évaluations ?

8. Résultats aux épreuves externes

- 8.1. Que faites-vous des résultats ? (après la rencontre au CO)
- 8.2. Remettez-vous en questions vos évaluations ? (barème, sévérité, types de questions, approche, gestion du temps ?)
- 8.3. Quel est l'importance des résultats des élèves ?
- 8.4. Quel devraient-ils être ? (plus importants, moins ?)
- 8.5. Vos élèves ont-ils les mêmes notes que lors des évaluations internes ? Pourquoi ?

9. Autres commentaires/compléments sur évaluation externe ? Autres remarques ? Il y a-t-il eu des questions qui vont ont marqué, étonné...

→ Si remarques ou questions par la suite, possibilité de me contacter.

Annexe B : Arborescence finale des catégories d'analyse

1. PROFIL GENERAL DE L ENSEIGNANT-----

- 1.1. age**
- 1.2. parcours professionnel**
- 1.3. ecole - village actuel**
- 1.4. classe actuelle**
- 1.5. autres**

2. PRATIQUES EVALUATION INTERNE-----

- 2.1. definition de l evaluation**
- 2.2. roles de l evaluation**
- 2.3. pratique evaluation generale**
- 2.4. planification de l evaluation dans un theme**
- 2.5. pratique d evaluation DIAGNOSTIC**
 - 2.5.1. defintion**
 - 2.5.2. roles buts**
 - 2.5.3. moments dans un theme**
 - 2.5.4. pratique - formes - moyens utilises**
 - 2.5.5. types de questions**
 - 2.5.6. niveau de difficulte - taxonomie**
 - 2.5.7. construction de l eval somm - documents de references**
- 2.6. pratique d evaluation FORMATIVE**
 - 2.6.1. defintion**
 - 2.6.2. roles buts**
 - 2.6.3. moments dans un theme**
 - 2.6.4. pratique - formes - moyens utilises**
 - 2.6.5. types de questions**
 - 2.6.6. niveau de difficulte - taxonomie**
 - 2.6.7. construction de l eval somm - documents de references**

2.7. pratique d evaluation SOMMATIVE

2.7.1. defintion

2.7.2. roles buts

2.7.3. moments dans un theme

2.7.4. pratique - formes - moyens utilises

2.7.5. types de questions

2.7.6. niveau de difficulte - taxonomie

2.7.7. construction de l eval somm - documents de references

2.7. autres

3. PERCEPTIONS DE L EVALUATION EXTERNE-----

3.1. experiences d evaluation externe

3.2. positionnement - arguments POUR

3.3. positionnement - arguments CONTRE

3.4. buts fonctions de l evaluation externe

3.4.1. pour les eleves

3.4.2. pour les enseignants

3.4.3. pour Etat

3.5. avantages - points positifs

3.5.1. pour les eleves

3.5.2. pour les enseignants

3.5.3. pour Etat

3.6. inconvenients - points negatifs

3.6.1. pour les eleves

3.6.2. pour les enseignants

3.6.3. pour Etat

3.7. critiques positives du CONTENU de l examen PPO

3.8. critiques negatives du CONTENU de l examen PPO

3.9. proposition d amelioration de la PPO

3.10. autres

4. EFFETS DE EVALUATION EXTERNE SUR LE CURRICULUM-----

- 4.1. planification de l'année
- 4.2. retrecissement du curriculum - perte de matière due à l'évaluation externe
- 4.3. réalisation de certaines activités spéciales extra scolaires
- 4.4. adéquation avec le programme
- 4.5. autres

5. EFFETS SUR LES PRATIQUES D'ENSEIGNEMENT-----

- 5.1. stratégies d'organisation - de préparation planification de l'année pour épreuve
- 5.2. matériel didactique
- 5.3. collaboration entre enseignants
- 5.4. techniques - exercices spécifiques de préparation à l'épreuve
- 5.5. changement de manières - de pratiques - d'enseignement
- 5.6. traitement des thèmes en profondeur
- 5.7. limites due à l'évaluation externes
- 5.8. innovations
- 5.9. rôle de l'enseignant
- 5.10. préparation de l'enseignant lui-même

6. EFFETS SUR L'EVALUATION INTERNE-----

- 6.1. évaluation interne des thèmes de l'évaluation externe
- 6.2. changements de pratiques d'évaluation interne dus à l'évaluation externe
- 6.3. sources - raisons des changements de pratiques
- 6.4. nouvelles idées - exemples - par l'évaluation externe
- 6.5. différences entre évaluation interne et externe
- 6.6. liens - concordance - entre évaluation interne et externe
- 6.7. ordre hiérarchique des types d'évaluation
- 6.8. autres

7. RESULTATS AUX EVALUATIONS EXTERNES-----

- 7.1. caracteristiques des eleves qui reussissent aux epreuves**
- 7.2. caracterisitques des eleves qui echouent aux epreuves**
- 7.3. utilisation des resultats**
- 7.4. sentiments a la reception des resultats**
- 7.5. imortance des resultats des eleves**
- 7.6. concordance entre resultats a l epreuve et notes a l annee**
- 7.7. changements de pratiques envisages dus aux resultats**
- 7.8. satisfactions des resultats**
- 7.9. insatisfactions des resultats**
- 7.10. part de l enseignant dans la reussite de l eleve**
- 7.11. autres**

8. AUTRES-----



Aux Responsables d'Etablissement

Madame, Monsieur,

L'Université de Fribourg, avec l'autorisation du SEnOF, mène une enquête qui a pour buts d'identifier les effets de l'évaluation externe (ex : épreuves cantonales, PISA...) sur les pratiques des enseignant-e-s primaires et d'en comprendre le fonctionnement, notamment au travers des conceptions de l'enseignement-apprentissage. Votre cercle scolaire a été sélectionné de manière aléatoire pour répondre à cette enquête.

Nous vous prions de bien vouloir faire remplir les **questionnaires** joints à cet envoi à toutes-tous les enseignant-e-s de 1P à 6P de votre cercle scolaire, puis de nous les retourner selon l'échéancier ci-dessous. Les enseignant-e-s de l'EE et d'AC en sont dispensé-e-s, n'étant concerné-e-s par aucune évaluation externe dans leur pratique.

Les réponses données à cette enquête seront traitées en toute **confidentialité et de manière anonyme** par l'Université de Fribourg. Un retour sur les résultats de cette enquête vous sera communiqué.

<p>Dès réception de cet envoi</p>	<p>Le RE distribue les questionnaires à toutes-tous les enseignant-e-s de son cercle scolaire (sauf EE et AC) et leur communique le délai du 15 février 2012.</p> <p><i>Nous vous proposons de distribuer ces questionnaires lors d'une réunion du mercredi après-midi (1^{er} février).</i></p>
<p>Jusqu'au 15 février 2012</p>	<p>Les enseignant-e-s remplissent le questionnaire (15 - 20 minutes) et le retournent au RE.</p> <p><i>Nous vous proposons de consacrer un moment pour compléter ces questionnaires lors d'une réunion du mercredi après-midi (1^{er} février).</i></p>
<p>Après la récolte des questionnaires</p>	<p>Le RE expédie les questionnaires complétés grâce à l'enveloppe affranchie, ou à l'adresse :</p> <p><i>Université de Fribourg Département des Sciences de l'éducation À l'attention de M. Gonzague Yerly Rue Faucigny 2 1700 Fribourg</i></p> <p><i>Pour toutes informations supplémentaires : 079 341 34 72</i></p>

D'ores et déjà, nous vous remercions de votre précieuse collaboration et vous adressons nos meilleures salutations.

Gonzague Yerly

Université de Fribourg



Chères-chers enseignant-e-s,

L'Université de Fribourg, avec l'autorisation de la DICS, mène une enquête qui a pour buts d'identifier les effets de l'évaluation externe (*ex : épreuves cantonales, PISA...*) sur les pratiques des enseignant-e-s primaires et d'en comprendre le fonctionnement.

Une **évaluation externe** est un dispositif d'évaluation des apprentissages des élèves qui n'est pas créé par l'enseignant-e mais par des personnes extérieures à la classe.

Vos réponses permettront d'identifier quelles sont les réactions des enseignant-e-s à ces épreuves et d'en comprendre les raisons, notamment au travers des conceptions de l'enseignement-apprentissage. C'est pourquoi nous vous prions de bien vouloir consacrer un peu de temps (15 à 20 minutes) pour répondre à nos questions, selon votre avis personnel. Répondez librement, les réponses seront traitées en toute **confidentialité** et de manière **anonyme** par l'Université de Fribourg.

Vous êtes priés de rendre ce questionnaire complété à votre Responsable d'Etablissement au plus tard le **mercredi 15 février 2012**.

**Vos réponses nous sont précieuses,
merci de votre collaboration!**

Merci de choisir dans la liste ci-dessous le type d'évaluation externe que vous rencontrez le plus souvent dans votre pratique. Pour les parties 1 et 2, veuillez répondre aux affirmations proposées en fonction de ce choix.

Dans ma pratique, je rencontre (ou ai rencontré) le plus souvent **le type d'évaluation externe** suivant :

- Les épreuves cantonales de cycle 1-2P
- Les épreuves cantonales de cycle 3-4P
- Les épreuves cantonales PPO de cycle 5-6P

Veillez mettre une croix dans la case correspondant à votre avis concernant les affirmations ci-dessous, sur l'échelle de 1 (= pas du tout d'accord) à 6 (= tout à fait d'accord).

Exemple :

<i>Je comprends comment fonctionne ce questionnaire.</i>	①	②	③	④	⑤	X
--	---	---	---	---	---	----------

N.B : Il est intentionnel pour la validité du questionnaire que plusieurs affirmations se ressemblent, n'y prêtez s.v.p. pas attention. Répondez de manière intuitive.

Partie 1 – Votre perception de l'évaluation externe sélectionnée		Pas du tout d'accord						Tout à fait d'accord					
		①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
1.1.	<i>Les exercices des EC* sont d'excellente qualité.</i>	①	②	③	④	⑤	⑥						
1.2.	<i>Les résultats de mes élèves aux EC me préoccupent beaucoup.</i>	①	②	③	④	⑤	⑥						
1.3.	<i>Les EC sont nécessaires pour compléter mes évaluations sommatives.</i>	①	②	③	④	⑤	⑥						
1.4.	<i>Les EC sont un exemple à suivre pour l'évaluation des apprentissages des élèves.</i>	①	②	③	④	⑤	⑥						
1.5.	<i>Les EC évaluent d'autres compétences que mes évaluations sommatives.</i>	①	②	③	④	⑤	⑥						
1.6.	<i>Les EC mettent une grosse pression sur les épaules des enseignants.</i>	①	②	③	④	⑤	⑥						
1.7.	<i>Les exercices des EC ressemblent fortement à ceux de mes propres évaluations sommatives.</i>	①	②	③	④	⑤	⑥						
1.8.	<i>Les EC n'évaluent pas les vraies capacités de mes élèves.</i>	①	②	③	④	⑤	⑥						
1.9.	<i>La réussite de mes élèves aux EC ne me stresse pas.</i>	①	②	③	④	⑤	⑥						
1.10.	<i>Les EC sont inutiles pour mon travail d'enseignant.</i>	①	②	③	④	⑤	⑥						
1.11.	<i>Les résultats de mes élèves aux EC sont les mêmes qu'à mes évaluations sommatives.</i>	①	②	③	④	⑤	⑥						
1.12.	<i>Je me sers des EC pour vérifier que mes pratiques d'enseignant sont adéquates.</i>	①	②	③	④	⑤	⑥						

* L'abréviation « **EC** » est utilisée pour « **épreuves cantonales** »

Partie 2 – Les effets de l'évaluation externe sur vos pratiques d'enseignement		Pas du tout d'accord				Tout à fait d'accord	
		①	②	③	④	⑤	⑥
2.1.	<i>Les EC me donnent régulièrement de nouvelles idées pour évaluer mes élèves.</i>	①	②	③	④	⑤	⑥
2.2.	<i>En vue des EC, je planifie mes leçons avec beaucoup plus de rigueur que d'habitude.</i>	①	②	③	④	⑤	⑥
2.3.	<i>Je ne fais pas d'exercices spécifiques pour préparer mes élèves aux EC.</i>	①	②	③	④	⑤	⑥
2.4.	<i>En vue de l'EC, j'observe encore plus les apprentissages de mes élèves.</i>	①	②	③	④	⑤	⑥
2.5.	<i>Sans l'exemple des évaluations cantonales, je n'évaluerais pas certaines compétences de mes élèves (par ex : compréhension orale).</i>	①	②	③	④	⑤	⑥
2.6.	<i>Dans ma classe, les branches évaluées aux EC n'ont pas la priorité sur les autres.</i>	①	②	③	④	⑤	⑥
2.7.	<i>Les EC m'incitent à collaborer davantage avec des collègues.</i>	①	②	③	④	⑤	⑥
2.8.	<i>J'organise d'importantes périodes de révisions pour préparer mes élèves aux EC.</i>	①	②	③	④	⑤	⑥
2.9.	<i>Je prends les EC comme exemples pour créer mes propres évaluations sommatives.</i>	①	②	③	④	⑤	⑥
2.10.	<i>Dans mes planifications, je privilégie les contenus qui seront évalués dans les EC au détriment d'autres branches.</i>	①	②	③	④	⑤	⑥
2.11.	<i>Je consacre une grande partie de mon temps d'enseignement à des exercices proches de ceux des EC pour préparer mes élèves.</i>	①	②	③	④	⑤	⑥
2.12.	<i>En vue des EC, je dois mener mon programme au pas de charge et n'ai pas de temps pour certains contenus.</i>	①	②	③	④	⑤	⑥
2.13.	<i>Je suis davantage attentif à la qualité de mon enseignement en vue des EC.</i>	①	②	③	④	⑤	⑥
2.14.	<i>Les EC n'influencent pas ma manière d'évaluer les apprentissages de mes élèves.</i>	①	②	③	④	⑤	⑥
2.15.	<i>Les EC m'amènent à réfléchir davantage à mes pratiques d'enseignant.</i>	①	②	③	④	⑤	⑥
2.16.	<i>Les EC font évoluer mes pratiques d'évaluation sommative.</i>	①	②	③	④	⑤	⑥
2.17.	<i>En raison des EC, je dois renoncer aux activités hors programme (par ex : camps, sorties,...).</i>	①	②	③	④	⑤	⑥
2.18.	<i>Les EC stimulent les discussions entre enseignants.</i>	①	②	③	④	⑤	⑥
2.19.	<i>Il est indispensable pour préparer les élèves de leur faire passer des EC des années précédentes.</i>	①	②	③	④	⑤	⑥
2.20.	<i>En vue des EC, je fais davantage de « drill » des contenus qui seront évalués.</i>	①	②	③	④	⑤	⑥
2.21.	<i>Certaines branches disparaissent en partie de ma grille horaire en vue des EC.</i>	①	②	③	④	⑤	⑥
2.22.	<i>J'échange davantage avec mes collègues en vue des EC.</i>	①	②	③	④	⑤	⑥

Partie 3 – Vos conceptions de l'apprentissage

		Pas du tout d'accord				Tout à fait d'accord	
		①	②	③	④	⑤	⑥
3.1.	<i>Pour bien apprendre, il faut être placé devant un obstacle à surmonter.</i>	①	②	③	④	⑤	⑥
3.2.	<i>On peut apprendre par la pratique ; pas toujours besoin de comprendre.</i>	①	②	③	④	⑤	⑥
3.3.	<i>Apprendre, c'est remplacer des conceptions antérieures par des conceptions nouvelles.</i>	①	②	③	④	⑤	⑥
3.4.	<i>La répétition n'est pas nécessaire à l'apprentissage.</i>	①	②	③	④	⑤	⑥
3.5.	<i>On apprend en mettant en pratique de façon répétitive.</i>	①	②	③	④	⑤	⑥
3.6.	<i>On pourrait dire que les prérequis sont les bases sur lesquelles on construit la suite des apprentissages.</i>	①	②	③	④	⑤	⑥
3.7.	<i>Pour apprendre, il faut être confronté à un problème que l'on cherche à résoudre.</i>	①	②	③	④	⑤	⑥
3.8.	<i>Pour apprendre, il faut d'abord abandonner ce qu'on savait auparavant.</i>	①	②	③	④	⑤	⑥
3.9.	<i>Il faut réaliser un certain nombre de fois le même type d'exercices pour apprendre.</i>	①	②	③	④	⑤	⑥
3.10.	<i>Apprendre, c'est établir des liens entre de nouvelles informations et des savoirs possédés antérieurement.</i>	①	②	③	④	⑤	⑥
3.11.	<i>On peut apprendre sans comprendre, par l'expérience directe des choses.</i>	①	②	③	④	⑤	⑥
3.12.	<i>Pour développer des compétences, l'enfant doit être placé face à des situations problèmes.</i>	①	②	③	④	⑤	⑥
3.13.	<i>Il n'y a pas toujours besoin de comprendre pour apprendre.</i>	①	②	③	④	⑤	⑥
3.14.	<i>Pour apprendre, il faut partir de ce que l'on sait déjà.</i>	①	②	③	④	⑤	⑥
3.15.	<i>Apprendre, c'est pouvoir remplacer une connaissance par une autre.</i>	①	②	③	④	⑤	⑥

Partie 4 – Votre perception de votre enseignement

		Pas du tout d'accord				Tout à fait d'accord	
		①	②	③	④	⑤	⑥
	<i>Je me sens capable de remédier efficacement aux problèmes de discipline en classe.</i>	①	②	③	④	⑤	⑥
4.2.	<i>Je me sens capable de motiver les élèves qui montrent peu d'intérêt pour les tâches scolaires.</i>	①	②	③	④	⑤	⑥
4.3.	<i>Lorsque les élèves arrivent à maîtriser aisément un nouveau concept, c'est parce que j'ai trouvé des moyens efficaces pour enseigner.</i>	①	②	③	④	⑤	⑥
4.4.	<i>Il m'est difficile d'évaluer les apprentissages de mes élèves avec une grande précision.</i>	①	②	③	④	⑤	⑥
4.5.	<i>Je me sens capable d'intéresser tous les élèves aux activités que j'organise.</i>	①	②	③	④	⑤	⑥
4.6.	<i>Après une séquence d'enseignement, je me sens capable de repérer les difficultés et les réussites de mes élèves.</i>	①	②	③	④	⑤	⑥
4.7.	<i>Je me sens capable d'amener tous les élèves à réussir une tâche même lorsqu'elle est difficile.</i>	①	②	③	④	⑤	⑥
4.8.	<i>Il m'est difficile de gérer les problèmes de comportement en classe.</i>	①	②	③	④	⑤	⑥
4.9.	<i>Même avec un bon enseignant, il y a toujours des élèves non motivés.</i>	①	②	③	④	⑤	⑥
4.10.	<i>Si un élève de ma classe est bruyant et dérange, je me sens capable de lui faire respecter les règles de la classe.</i>	①	②	③	④	⑤	⑥
4.11.	<i>Je me sens capable de déterminer de manière précise si un élève maîtrise ou non un contenu.</i>	①	②	③	④	⑤	⑥
4.12.	<i>Je me sens capable d'amener tous les élèves à la maîtrise des savoirs enseignés à l'école.</i>	①	②	③	④	⑤	⑥

Partie 5 – Vos conceptions de l'évaluation en général

		Pas du tout d'accord				Tout à fait d'accord	
		①	②	③	④	⑤	⑥
<i>N.B. : Les affirmations suivantes portent sur l'évaluation en général. Il ne s'agit plus des EC.</i>							
5.1.	<i>L'évaluation permet de cerner les progrès que les élèves ont réalisés.</i>	①	②	③	④	⑤	⑥
5.2.	<i>Les résultats des évaluations permettent d'identifier les élèves qui n'ont pas d'aptitudes scolaires.</i>	①	②	③	④	⑤	⑥
5.3.	<i>Les évaluations permettent de repérer les élèves qui ont besoin d'une explication supplémentaire.</i>	①	②	③	④	⑤	⑥
5.4.	<i>Les évaluations sanctionnent les élèves qui n'ont pas travaillé.</i>	①	②	③	④	⑤	⑥
5.5.	<i>Les évaluations permettent de repérer les bons et les mauvais élèves.</i>	①	②	③	④	⑤	⑥
5.6.	<i>L'analyse des résultats des élèves aux évaluations permet d'ajuster l'enseignement en fonction de leurs forces et faiblesses.</i>	①	②	③	④	⑤	⑥
5.7.	<i>Les évaluations sont l'occasion pour les élèves de montrer ce qu'ils valent.</i>	①	②	③	④	⑤	⑥
5.8.	<i>Les évaluations permettent de repérer ce qu'il faut retravailler avec mes élèves.</i>	①	②	③	④	⑤	⑥

Partie 6 – Informations personnelles

6.1. Quel est votre sexe ?	<input type="checkbox"/> femme <input type="checkbox"/> homme										
6.2. Quel âge avez-vous ? ans										
6.3. Quel est votre pourcentage d'activité dans l'enseignement primaire? %										
6.4. Dans quel(s) degré(s) enseignez-vous actuellement ? (plusieurs croix possibles)	<input type="checkbox"/> 1P <input type="checkbox"/> 2P <input type="checkbox"/> 3P <input type="checkbox"/> 4P <input type="checkbox"/> 5P <input type="checkbox"/> 6P										
6.5. Combien d'années d'expérience avez-vous dans l'enseignement primaire ?	<input type="checkbox"/> 0 – 2 ans <input type="checkbox"/> 11 – 20 ans <input type="checkbox"/> 3 – 4 ans <input type="checkbox"/> 21 – 30 ans <input type="checkbox"/> 5 – 10 ans <input type="checkbox"/> + de 30 ans										
6.6. Dans quel(s) cycle(s) avez-vous été titulaire d'une classe dans votre carrière ? (plusieurs croix possible)	<input type="checkbox"/> 1 – 2 P <input type="checkbox"/> CO <input type="checkbox"/> 3 – 4 P <input type="checkbox"/> Ens. spécialisé/MCDI <input type="checkbox"/> 5 – 6 P <input type="checkbox"/> Autre :										
6.7. Combien d'années avez-vous enseigné dans le cycle où vous enseignez actuellement ?	<input type="checkbox"/> 0 – 2 ans <input type="checkbox"/> 11 – 20 ans <input type="checkbox"/> 3 – 4 ans <input type="checkbox"/> 21 – 30 ans <input type="checkbox"/> 5 – 10 ans <input type="checkbox"/> + de 30 ans										
6.8. Dans quel type d'école enseignez-vous ?	<input type="checkbox"/> plusieurs classes du même degré dans le même bâtiment <input type="checkbox"/> une seule classe du même degré dans le même bâtiment										
6.9. Dans quel type de commune enseignez-vous?	<input type="checkbox"/> campagne <input type="checkbox"/> périurbaine <input type="checkbox"/> ville										
6.10. Comment définiriez-vous le niveau de votre classe actuelle ?	<table style="width: 100%; text-align: center;"> <tr> <td>excellent</td> <td>bon</td> <td>moyen</td> <td>faible</td> <td>très faible</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> </table>	excellent	bon	moyen	faible	très faible	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
excellent	bon	moyen	faible	très faible							
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>							
6.11. Selon vous, quel est votre degré de collaboration avec vos collègues ? (autres que la personne avec qui vous formez év. un duo)	<table style="width: 100%; text-align: center;"> <tr> <td>très souvent</td> <td>assez souvent</td> <td>parfois</td> <td>rarement</td> <td>jamais</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> </table>	très souvent	assez souvent	parfois	rarement	jamais	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
très souvent	assez souvent	parfois	rarement	jamais							
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>							
6.12. Sous quelle forme collaborez-vous le plus souvent ? (autres que la personne avec qui vous formez év. un duo)	<input type="checkbox"/> Je collabore le plus souvent avec des collègues du même cycle, mais de différents cercles scolaires. <input type="checkbox"/> Je collabore le plus souvent avec des collègues du même cycle et du même cercle scolaire. <input type="checkbox"/> Autre :										
6.13. Quel type de formation avez-vous suivi pour devenir enseignant(e) ?	<input type="checkbox"/> Ecole normale <input type="checkbox"/> HEP <input type="checkbox"/> Autre :										
6.14. Avez-vous suivi des formations complémentaires ?	<input type="checkbox"/> Formation continue sur <u>l'évaluation</u> <input type="checkbox"/> Formation de Responsable d'établissement <input type="checkbox"/> Licence ou Master en										
6.15. Avez-vous déjà participé à la conception d'EC ?	<input type="checkbox"/> oui <input type="checkbox"/> non										

Commentaires éventuels :

MERCI DE VOTRE PARTICIPATION !

Un retour sur les résultats vous sera communiqué par le biais de votre Responsable d'Etablissement.

Annexe E : Corrélations entre les croyances des enseignants à propos de l'enseignement

		Corrélations									
		Déséquilibre	Répétition	Apprentissage implicite	Eradication des connaissances antérieures	Intégration des connaissances antérieures	Maîtrise des savoirs	Gestion de la discipline	Evaluation des apprentissages	Evaluation formative	Evaluation normative
Déséquilibre	Corrélation de Pearson	1	.066	.040	.169**	.234**	.023	-.054	.043	.122*	-.002
	Sig. (bilatérale)		.246	.483	.003	.000	.691	.342	.450	.031	.975
	N	311	311	311	311	311	309	309	309	311	311
Répétition	Corrélation de Pearson	.066	1	.148**	-.103	.209**	.037	-.016	.001	-.030	.261**
	Sig. (bilatérale)	.246		.009	.070	.000	.513	.775	.993	.592	.000
	N	311	311	311	311	311	309	309	309	311	311
Apprentissage implicite	Corrélation de Pearson	.040	.148**	1	.141*	-.156**	-.048	-.095	.021	.104	.049
	Sig. (bilatérale)	.483	.009		.013	.006	.398	.095	.717	.068	.386
	N	311	311	311	311	311	309	309	309	311	311
Eradication des connaissances antérieures	Corrélation de Pearson	.169**	-.103	.141*	1	-.213**	-.039	-.073	-.022	.066	.080
	Sig. (bilatérale)	.003	.070	.013		.000	.499	.201	.694	.247	.157
	N	311	311	311	311	311	309	309	309	311	311
Intégration des connaissances antérieures	Corrélation de Pearson	.234**	.209**	-.156**	-.213**	1	.152**	.140*	.095	.029	.002
	Sig. (bilatérale)	.000	.000	.006	.000		.007	.014	.097	.610	.971
	N	311	311	311	311	311	309	309	309	311	311
Maîtrise des savoirs	Corrélation de Pearson	.023	.037	-.048	-.039	.152**	1	.304**	.378**	.010	.039
	Sig. (bilatérale)	.691	.513	.398	.499	.007		.000	.000	.859	.499
	N	309	309	309	309	309	309	309	309	309	309
Gestion de la discipline	Corrélation de Pearson	-.054	-.016	-.095	-.073	.140*	.304**	1	.295**	.034	-.003
	Sig. (bilatérale)	.342	.775	.095	.201	.014	.000		.000	.548	.955
	N	309	309	309	309	309	309	309	309	309	309
Evaluation des apprentissages	Corrélation de Pearson	.043	.001	.021	-.022	.095	.378**	.295**	1	.155**	.052
	Sig. (bilatérale)	.450	.993	.717	.694	.097	.000	.000		.006	.361
	N	309	309	309	309	309	309	309	309	309	309
Evaluation formative	Corrélation de Pearson	.122*	-.030	.104	.066	.029	.010	.034	.155**	1	.241**
	Sig. (bilatérale)	.031	.592	.068	.247	.610	.859	.548	.006		.000
	N	311	311	311	311	311	309	309	309	311	311
Evaluation normative	Corrélation de Pearson	-.002	.261**	.049	.080	.002	.039	-.003	.052	.241**	1
	Sig. (bilatérale)	.975	.000	.386	.157	.971	.499	.955	.361	.000	
	N	311	311	311	311	311	309	309	309	311	311

** La corrélation est significative au niveau 0.01 (bilatéral).

* La corrélation est significative au niveau 0.05 (bilatéral).

Gonzague YERLY

08.05.1984

Formation

- 2009 - 2014 **Doctorat ès Lettres en Sciences de l'éducation**
Université de Fribourg (CH)
- 2007-2009 **Master of Science en Sciences de l'éducation**
Option « Innovation et formation »
Université de Fribourg (CH)
- 2004 - 2007 **Bachelor of Arts in Pre-Primary and Primary Education**
Diplôme d'enseignement primaire aux degrés préscolaires et primaires
Haute Ecole Pédagogique de Fribourg (CH)
- 1999 - 2003 **Maturité fédérale**
Option Economie et Droit
Collège Saint-Michel, Fribourg (CH)

Expériences professionnelles

Recherche et enseignement universitaire

- 2009 - 2014 **Assistant diplômé (50%)**, recherche et enseignement
Département des Sciences de l'éducation, Université de Fribourg (CH)
- 2008 - 2009 **Sous-assistant**, recherche et enseignement
Département des Sciences de l'éducation, Université de Fribourg (CH)

Formation d'adultes

- 2008 **Collaborateur spécifique spécialisé**, coordinateur de projet
Commission cantonale de la formation des adultes du canton de Fribourg (CH)

Enseignement au degré primaire

- 2009 - 2014 **Enseignant (50%)**, classes de 5P-6P
Direction de l'Instruction publique du canton de Fribourg (CH)

Annexe G : Déclaration sur l'honneur

Je déclare sur mon honneur que ma thèse est une œuvre personnelle, composée sans concours extérieur non autorisé, et qu'elle n'a pas été présentée devant une autre Faculté.



Gonzague Yerly