

Développement d'un modèle d'évaluation adapté au PER

Rapport scientifique du projet d'épreuves
romandes communes

Viridiana Marc
Martine Wirthner
avec la collaboration de Sabrina Uldry



Institut de recherche et de documentation pédagogique

Développement d'un modèle d'évaluation adapté au PER

Rapport scientifique du projet d'épreuves
romandes communes

Viridiana Marc

Martine Wirthner

avec la collaboration de Sabrina Uldry

Remerciements

La partie scientifique du projet EpRoCom a été traitée par deux groupes de travail, l'un composé de spécialistes de la didactique des mathématiques, l'autre de spécialistes de la didactique du français. Les travaux de ces groupes ont été denses et riches et nous tenons à dire notre reconnaissance à chacune et à chacun de leurs membres; leurs noms figurent à la fin de ce rapport.

Un tel travail n'aurait pas pu être réalisé sans un suivi attentif des dossiers; nous avons ainsi bénéficié de la précieuse collaboration de Caroline Sperisen (suivi administratif et travail sur la classification des tâches), de Bertrand Vermot (suivi informatique des données, participation aux travaux des groupes), de Sabrina Uldry (étude sur la classification des tâches), assistants; et d'Amélie Guay, stagiaire, pour une étude sur les consignes des épreuves.

Plusieurs collaborateurs scientifiques de l'IRDp ont été sollicités à titre d'experts et ont ainsi sensiblement contribué à l'avancement de nos travaux: Luc-Olivier Pochon pour son expérience sur la classification des tâches (ontologies); Alina Matei pour la partie statistique de la recherche; Jean-François de Pietro, Murielle Roth, ainsi que Diego Corti, appelés à réagir aux résultats obtenus, en particulier – pour les deux premiers – dans la discipline français. Que tous soient vivement remerciés.

Rencontrés à notre demande lors du 24^{ème} colloque de l'ADMEE à Luxembourg en 2012, les spécialistes de l'évaluation de compétences que sont Bernard Rey, Sabine Kahn, Xavier Roegiers et Christian Dierendonck nous ont accordé de leur temps et de leur attention pour réagir à notre projet. Tout comme leurs travaux, leurs avis nous ont été précieux. Nous tenons également à leur dire notre reconnaissance.

Il ne peut y avoir publication sans relecture et mise en page du texte. Toutes trois de l'IRDp, Véronique Françoise et Françoise Landry ont assuré la relecture de l'ensemble du rapport, alors que Doris Penot a finalisé sa mise en page. Nous leur disons notre gratitude, conscientes de la minutie et de la patience nécessaires et dont elles ont fait preuve pour accomplir de telles tâches.

Publication	Institut de recherche et de documentation pédagogique – IRDP Neuchâtel
Mise en page	Doris Penot (IRDp)
Couverture	Marc-Olivier Schatz
Photo	Viridiana Marc
Diffusion	IRDp Case postale 556 CH-2002 Neuchâtel www.irdp.ch documentation@irdp.ch

Table des matières

Note préalable	5
Synthèse	7
1. CADRE GÉNÉRAL	9
1.1 Contexte romand	9
1.2 Mandat et organisation du projet	10
1.3 Panorama des évaluations cantonales	12
1.3.1 Évaluations cantonales au cycle 1	13
1.3.2 Évaluations cantonales au cycle 2	14
1.3.3 Évaluations cantonales au cycle 3	14
1.4 Interprétations de l'article 15 de la Convention scolaire romande (CSR) ...	14
1.5 Scénarios explorés	16
1.5.1 Scénario 1 : évaluation sommative	16
1.5.2 Scénario 2 : évaluation diagnostique	18
2. RÉFLEXIONS GÉNÉRALES.....	21
2.1 L'évolution des référentiels d'enseignement	21
2.2 L'évaluation de compétences	25
2.3 Les buts et visées de l'évaluation	29
3. DÉVELOPPEMENT D'UN MODÈLE D'ÉVALUATION ADAPTÉ AU PER ..	31
3.1 Un référentiel : le PER	31
3.2 Exemples d'évaluation en Communauté française de Belgique	37
3.3 Éléments pour une évaluation adaptée au PER.....	40
3.3.1 Un modèle d'évaluation en deux parties complémentaires	40
3.3.2 Un développement possible : une épreuve adaptative	43
3.3.3 Un modèle adaptable à la visée évaluative	47
4. ÉLABORATION D'UN PROTOTYPE D'ÉVALUATION EN DEUX PHASES ...	51
4.1 Les options retenues dans l'élaboration des épreuves.....	52
4.2 Évaluation en deux phases en français à la fin de cycle 1	53
4.2.1 Caractéristiques de chaque phase pour le français	53
4.2.2 Quelques choix dans l'élaboration de l'épreuve adaptative.....	54
4.2.3 Quelques choix dans l'élaboration de l'épreuve papier.....	59
4.3 Évaluation en deux phases en mathématiques fin de cycle 1	65

4.3.1	<i>Répertoire des éléments du référentiel à évaluer</i>	65
4.3.2	<i>Quelques choix dans l'élaboration de l'épreuve adaptative</i>	66
4.3.3	<i>Quelques choix dans l'élaboration de l'épreuve papier</i>	70
4.4	Recueil des données	71
4.4.1	<i>Gestion en ligne et récupération automatique des informations</i>	72
4.4.2	<i>Recueil des données des épreuves papier</i>	74
4.5	Avis des enseignants	75
4.5.1	<i>A propos des entretiens pré-passation</i>	76
4.5.2	<i>A propos des entretiens post-passation</i>	77
4.5.3	<i>Avis des enseignants sur l'utilité diagnostique des épreuves</i>	79
4.5.4	<i>Améliorations proposées par les enseignants</i>	79
4.6	Analyse des consignes	80
5.	ANALYSE DES DONNÉES DES ÉVALUATIONS EN DEUX PHASES	85
5.1	Quelques remarques relatives à notre modèle	85
5.2	Analyses et résultats	87
5.2.1	<i>Méthodologie pour l'analyse de données</i>	89
5.2.2	<i>Les résultats des épreuves de français – cycle 1</i>	92
5.2.3	<i>Les résultats de l'épreuve de mathématiques – cycle 1</i>	114
5.2.4	<i>Les liens entre l'épreuve français et l'épreuve mathématiques</i>	133
5.3	Quels messages pour quels publics ?	139
6.	CONCLUSIONS	141
7.	LES PERSPECTIVES	147
7.1	Les conditions d'élaboration d'une évaluation	147
7.1.1	<i>Pistes pour l'élaboration d'épreuves</i>	148
7.1.2	<i>Conditions à assurer pour une épreuve informatisée</i>	149
7.2	Les conditions d'élaboration de tâches	150
7.2.1	<i>Pistes pour la rédaction de tâches</i>	150
7.2.2	<i>Éléments de classification</i>	151
7.2.3	<i>Vers une banque d'items régionale</i>	153
8.	ANNEXES	155
8.1	Liste des membres participants	155
8.2	Description des évaluations cantonales par cycle	156
8.3	Description de l'évaluation en deux phases en français	159

8.3.1	<i>Description des tâches composant l'EA en français</i>	159
8.3.2	<i>Description des tâches composant l'EA en français</i>	160
8.3.3	<i>Description de l'épreuve papier (EP) en français</i>	170
8.3.4	<i>Description des tâches composant l'EP en français</i>	171
8.4	Description de l'évaluation en deux phases en mathématiques	176
8.4.1	<i>Description de l'épreuve adaptative (EA) en mathématiques</i>	176
8.4.2	<i>Description des tâches composant l'EA en mathématiques</i>	177
8.4.3	<i>Description de l'épreuve papier (EP) en mathématiques</i>	188
8.4.4	<i>Description des tâches composant l'EP en mathématiques</i>	188
BIBLIOGRAPHIE		193
GLOSSAIRE		197

Note préalable

Le présent rapport est destiné à faire connaître les travaux de recherche menés corrélativement au projet d'épreuves romandes communes (EpRoCom) dans le cadre d'un mandat de la CIIP¹ confié à l'IRD. Un premier rapport, intitulé Épreuves romandes communes: de l'analyse des épreuves cantonales à un modèle d'évaluation adapté au PER² a déjà été livré aux mandants de ce projet en juin 2011.

Parce que nous inscrivons nos travaux dans la perspective d'une recherche appliquée, nous présenterons essentiellement ici nos réflexions et choix au regard, bien sûr, de travaux de recherche, mais également de contraintes pragmatiques. Nous renvoyons donc les lecteurs intéressés aux références bibliographiques permettant d'approfondir les réflexions théoriques proposées dans ce texte.

Les analyses effectuées dans le cadre de cette recherche ne présupposent en rien de la forme et des visées que revêtiront les épreuves romandes mais constituent une base commune de développement et de réflexion sur l'évaluation des objectifs du PER. En ce sens, ce rapport s'adresse plus particulièrement aux responsables cantonaux en charge de l'évaluation, mais peut également intéresser les formateurs et enseignants susceptibles de se pencher sur l'élaboration de travaux d'évaluation pour la classe, le PER étant pour l'instant le seul référentiel à disposition, bien que non conçu comme référentiel d'évaluation.

La notation des années scolaires se réfère à celle utilisée dans le PER, en cohérence avec le Concordat HarmoS, parlant des années 1 à 11 que compte désormais la scolarité obligatoire.

Les mentions des personnes dans ce document valent indifféremment pour le masculin et le féminin.

¹ CIIP SR-TI : Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin

² PER : Plan d'études romand

Synthèse

Dans l'article 15 de la Convention scolaire romande (CSR), la CIIP (Conférence inter-cantonale de l'instruction publique de la Suisse romande et du Tessin) exprime sa volonté de développer des épreuves romandes communes. Le contenu de cet article soulève toutefois la question des finalités de telles épreuves. Le présent rapport, sans y apporter directement une réponse, met en avant les différentes réflexions menées parallèlement au projet EpRoCom (Épreuves romandes communes) et dont le but premier est de caractériser une évaluation permettant la mesure des objectifs du PER (Plan d'études romand).

Nous présentons tout d'abord les situations cantonales en matière d'évaluation ainsi que le cadre du mandat. Nous nous penchons ensuite sur des éléments théoriques permettant d'étayer les choix effectués dans nos développements ainsi que sur un modèle évaluatif susceptible de répondre aux exigences d'une évaluation adaptée au PER.

Enfin, nous décrivons une épreuve-prototype diagnostique (en mathématiques et en français) permettant une première mise à l'essai du modèle évaluatif proposé. Ceci nous amène à envisager des éléments de classification des items et des tâches complexes, notamment dans la perspective de développer une banque d'items, ainsi que le traitement possible des données et les améliorations souhaitées.

Zusammenfassung

Im Artikel 15 der *Convention scolaire romande* (CSR) hält die Westschweizer Erziehungsdirektorenkonferenz (CIIP) fest, dass sie gemeinsame Westschweizer Prüfungen zu entwickeln beabsichtigt. Der Inhalt dieses Artikels lässt jedoch die Frage offen, welches die Ziele solcher Prüfungen sind. Der vorliegende Bericht kann darauf zwar keine direkte Antwort geben, doch er bilanziert die verschiedenen Überlegungen, die parallel zum Projekt *EpRoCom* gemacht worden sind und deren oberstes Ziel es ist, eine Form der Evaluierung zu skizzieren, mit Hilfe derer die Zielsetzungen des Westschweizer Lehrplans (PER) messbar gemacht werden können.

Zunächst wird die Situation in den Westschweizer Kantonen im Bereich Evaluation sowie das Mandat vorgestellt. Anschliessend wird die Theorie, welche unser Modell untermauert, erläutert, sowie ein Evaluationsmodell, das den Anforderungen für eine lehrplangerechte Evaluierung genügen soll.

Schliesslich wird auch eine prototypische diagnostische Prüfung beschrieben, mit der das vorgeschlagene Modell erstmals getestet werden kann. Dies erlaubt es, Grundlagen für eine Klassifizierung der Items und für komplexe Aufgaben zusammenzustellen (im Hinblick auf die Erstellung einer Item-Datenbank) sowie die Möglichkeiten für die Datenbearbeitung und die gewünschten Verbesserungen aufzuzeigen.

Summary

In article 15 of the *Convention scolaire romande* (CSR), the CIIP expresses the wish to develop common tests for the French-speaking cantons. However, the content of the legal text raises questions about the purposes of such tests. This report cannot give

definite answers, but it shows the considerations that have been made while working on the project *EpRoCom*, which aim at characterising an evaluation capable of measuring the goals of the regional syllabus (*Plan d'études romand – PER*).

First, we present the situation in terms of evaluation in the French-speaking cantons, as well as the framework of the mandate. Second, we focus on the one hand on theoretical considerations that permit us to found the choices made during our development and on the other hand on an evaluative model that is capable of meeting the requirements of an evaluation adjusted to the PER.

Finally, we describe a prototype for a diagnostic test that serves as a first attempt to concretise the proposed evaluative model. Thereby, we can consider the principles of classification for items and complex tasks, especially considering the goal to establish an item database, as well as the data manipulations that are rendered possible and the desired amendments.

Sintesi

Nell'articolo 15 della Convenzione scolastica romanda (CSR), la CIIP esprime la volontà di sviluppare delle prove romande comuni. Il contenuto di questo articolo solleva tuttavia la questione delle finalità di tali prove. Questo rapporto, senza potervi rispondere direttamente, presenta le diverse riflessioni condotte parallelamente al progetto *EpRoCom*, il cui scopo principale è di caratterizzare una valutazione che permetta di misurare gli obiettivi del PER (piano di studi romando).

Dapprima, noi presentiamo le situazioni della valutazione a livello cantonale e il mandato. In seguito, ci occuperemo degli elementi teorici che permettono di spiegare le scelte effettuate ed il modello di valutazione in grado di rispondere alle esigenze adattate al PER.

Infine, viene descritto un prototipo di prova diagnostica per una prima verifica del suddetto modello valutativo. Questo ci permette di immaginare degli elementi per la classificazione degli items e dei compiti complessi, soprattutto nella prospettiva di sviluppare una banca dati, così come nel possibile trattamento di questi dati e dei successivi miglioramenti voluti.

1. Cadre général

1.1 Contexte romand

Dès 2003, les travaux romands (CIIP) d'écriture d'un plan d'études cadre et, en 2004, les conclusions du rapport du GRETEL³ (*Lignes directrices pour un système d'évaluation du travail des élèves dans la scolarité obligatoire*) offrent respectivement les prémices d'un plan d'études et une clarification des divers types d'évaluation présents en Suisse romande. Axé sur l'évaluation du travail des élèves, le texte final du GRETEL distingue ce qui relève :

- du travail de l'élève (évaluation en classe, formative, diagnostique ou sommative);
- de la mesure de l'atteinte des objectifs fixés par le référentiel (évaluations externes cantonales ou régionales, sommatives ou bilan, en vue d'une régulation de l'enseignement).

À cela s'ajoutent encore les travaux de monitoring HarmoS. Ils sont destinés à établir des standards nationaux à partir desquels constituer une mesure de l'efficacité du système scolaire suisse (sur échantillons d'élèves des différentes régions).

Ces travaux préparatoires débouchent sur des propositions reprises, en 2006, dans les mesures CIIP accompagnant la fin de la rédaction du Plan d'études cadre. La Mesure 5 notamment, *Élaboration des instruments d'évaluation*, évoque la nécessité de développer des outils de test permettant l'appréciation de l'implantation du curriculum. Elle se concrétise dans l'article 15 de la Convention scolaire romande de la manière suivante :

Article 15 – Épreuves romandes

¹ La CIIP organise des épreuves romandes communes à l'espace romand de la formation en vue de vérifier l'atteinte des objectifs du plan d'études.

² En fin de cycle ou à la fin du secondaire I, si la discipline choisie pour l'épreuve romande commune correspond à celle d'un test de référence vérifiant un standard national, le test de référence peut servir d'épreuve commune.

Dès 2004, à la suite des travaux du GRETEL, l'IRDП lance une première réflexion sur les conditions de création d'épreuves romandes avec des groupes de spécialistes issus des cantons.

³ Groupe de référence pour l'évaluation du travail des élèves.

Un panorama des évaluations cantonales est élaboré pour comprendre la situation de l'évaluation en Suisse romande. Le tableau ci-contre met d'emblée en évidence le nombre important des évaluations dans les cantons. L'analyse de la visée de chacune s'avère nécessaire pour en saisir les enjeux (cf. point 1.3).

Tableau 1.1 : Panorama des évaluations cantonales

	Évaluations cantonales								
	3e	4e	5e	6e	7e	8e	9e	10e	11e
	1P	2P	3P	4P	5P	6P	7e	8e	9e
BE						X			
FR		X		X		X	X	X	X
GE		X		X		X	X	X	X
JU				X		X		X	
NE	X	X	X	X	X	X		X	
VS		X	X	X	X	X		X	X
VD		X		X		X		X	X

Au terme de plusieurs rencontres, un rapport (Pochon & Wirthner, 2007) est élaboré, mettant en évidence la nécessité de créer des collaborations intercantionales en vue de constituer un pôle romand de compétences en matière d'évaluation (notamment pour la construction d'une méthodologie commune, la constitution d'une banque d'items, ...).

Par ailleurs, sous l'impulsion du canton de Berne, des travaux d'écriture d'un plan d'études commun débutent en 2005, associant les cantons de Neuchâtel, du Jura ainsi que de Fribourg. Ralliant peu à peu les autres cantons, ce projet devient romand, le 20 septembre 2007, pour aboutir au Plan d'études romand (PER). Celui-ci, soumis à des experts, fait ensuite l'objet d'une consultation en automne 2008. Des travaux d'amélioration sont entrepris en automne 2009 pour parvenir à une version finale du PER au printemps 2010. La mise en application de ce plan d'études est prévue dans les classes dès août 2011, avec des variations selon les cantons et les années scolaires. Ceux-ci ont également la charge de réorganiser leurs structures scolaires en conformité avec le Concordat HarmoS.

Entretemps, le 21 juin 2007, la CIIP propose la Convention scolaire romande (CSR). Elle entre dans une procédure de ratification par les parlements cantonaux; les cantons la ratifient un à un jusqu'à son entrée en vigueur en 2009.

1.2 Mandat et organisation du projet

En septembre 2007, la CIIP attribue un mandat à l'IRDP pour que soit étudiée la possibilité de créer des épreuves romandes en référence au PER, dans les disciplines scolaires français et mathématiques. Préparé en détail par l'IRDP, le projet est d'abord accepté le 21 avril 2008 par le Comité de pilotage puis par la plénière de la CIIP lors de sa séance des 5 et 6 juin 2008.

Le mandat de recherche (Marc & Wirthner, 2012) et de développement d'épreuves romandes communes (EpRoCom), prévu sur trois ans, est donc donné formellement à l'IRDP au printemps 2008, prenant appui d'une part sur les travaux du PER et d'autre part sur les travaux préalablement entrepris à l'IRDP.

L'équipe de projet en charge du dossier⁴ met sur pied, pour la rentrée scolaire 2008 et jusqu'à fin 2010, d'une part un *Consortium romand* et, d'autre part, des *Groupes de travail* (un en français et un en mathématiques) pour l'élaboration d'épreuves (voir liste à l'annexe 8.1).

Le Consortium romand

Le *Consortium romand*, est constitué de représentants cantonaux en charge du dossier évaluation. Un membre des *Groupes de travail* fait partie de ce consortium. S'appuyant sur les diverses évaluations cantonales, il est chargé de déterminer un concept d'évaluation commun pour des épreuves romandes ainsi que d'accompagner les groupes de travail dans leur élaboration d'épreuves.

Les échanges sur les différents dispositifs d'évaluation cantonaux vont très vite montrer une grande disparité entre eux (comme le montre le tableau 1.1), rendant impossible pour les membres du *Consortium* de décider eux seuls quelle finalité donner aux épreuves romandes. Plusieurs scénarios seront alors élaborés (voir point 1.5) et discutés afin de clarifier, pour les décideurs politiques, les enjeux d'épreuves romandes en fonction de leur finalité (voir point 2.3). Dès le printemps 2009, la *Conférence des secrétaires généraux* décide de mettre en suspens les travaux du *Consortium romand* pour permettre une prise en charge de ces réflexions par les chefs de département eux-mêmes. En revanche, le travail de développement d'épreuve par les *Groupes de travail* se poursuit jusqu'en 2011.

Les Groupes de travail

Les *Groupes de travail* sont constitués de spécialistes de l'évaluation de Suisse romande. Il s'agit autant de chercheurs en évaluation que de didacticiens, formateurs HEP et enseignants concepteurs d'épreuves dans leur canton.

Partant du PER qui propose un référentiel structuré identiquement pour chaque discipline, les travaux des deux groupes se font à la fois en commun et en parallèle. La volonté est de conserver ainsi une cohérence entre les deux disciplines et, dans la mesure du possible, non seulement de constituer un modèle d'évaluation semblable mais également de s'entendre, plus concrètement, sur des modalités communes d'élaboration de tâches.

Après l'étude des diverses évaluations cantonales, les groupes travaillent à stabiliser les définitions des types d'épreuves ainsi qu'à explorer et à préparer des scénarios qui seront repris par le consortium (voir point 1.5). Les groupes prennent également le parti d'imaginer un modèle évaluatif qui tienne compte des différents éléments présents dans le PER (voir chapitre 3) et d'explorer concrètement ce qu'implique une interprétation de ses contenus pour un objectif d'apprentissage donné, en particulier au cycle 1 (voir chapitre 4).

Les groupes travaillent largement à l'élaboration et à l'ajustement de questions mesurant les éléments du PER. La question de la classification à assurer pour la mise en place d'une banque d'items est également discutée (voir point 7.2.2). Les analyses des

⁴ Une cheffe de projet est désignée en la personne de Viridiana Marc.

liens entre les diverses questions doivent leur permettre de valider le modèle d'évaluation exploré.

1.3 Panorama des évaluations cantonales

S'inspirant de différentes études (voir chapitre 2), les *Groupes de travail* et le *Consortium romand* stabilisent des définitions pour les divers types d'évaluation présents dans le paysage suisse. Notamment les définitions d'évaluation à visée sommative, certificative, diagnostique, ou encore d'évaluation de sélection, d'orientation ou bilan ont été arrêtées afin de se repérer dans les divers types d'évaluations cantonales et de s'accorder sur l'emploi d'une terminologie commune. Le terme « test de référence » reste réservé à la mesure des standards HarmoS (voir glossaire).

Leur effort s'est plus particulièrement concentré sur deux grands types d'évaluation dont les finalités coexistent dans les cantons romands et ont des implications fortes sur la constitution des épreuves qui les sous-tendent :

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. »

Ces deux définitions, inspirées des travaux du GRETEL et stabilisées par les groupes et le consortium, permettent de rendre compte de l'ensemble des situations cantonales et de mieux comprendre où se situent les enjeux évaluatifs, notamment de cerner ce qu'il s'agirait de prendre en considération pour d'éventuelles épreuves romandes.

Au regard de ces deux définitions, la situation romande, pour les années du primaire et du secondaire I, se présente comme suit :

Tableau 1.3 : Panorama des évaluations cantonales en fonction de leurs visées sommative et diagnostique (état en 2012).

	Évaluations à visée sommative										Évaluations à visée diagnostique									
	3e	4e	5e	6e	7e	8e	9e	10e	11e	3e	4e	5e	6e	7e	8e	9e	10e	11e		
	1P	2P	3P	4P	5P	6P	7e	8e	9e	1P	2P	3P	4P	5P	6P	7e	8e	9e		
BE															X					
FR						X			X		X		X			X	X			
GE		X		X		X	X	X	X											
JU						X							X					X		
NE						X		X		X	X	X	X	X				X ^a		
VS				X		X		X	X		X ^a	X ^a	X ^a	X ^a	X ^a			X ^a		
VD		X		X		X		X	X											
		2		3		6	1	3	4	1	3	2	4	2	2	1	4	1		

□ Épreuve inscrite dans la loi scolaire du canton ^a: Épreuves à disposition des enseignants, passation libre

Quelques constats :

- Certains cantons ont une culture forte de l'évaluation de type sommatif et très peu de l'évaluation de type diagnostique (comme GE et VD), alors que d'autres ont une culture très légère de l'évaluation externe qui, si elle existe, aura plutôt tendance à être de type diagnostique (comme BE et JU).
- Certaines années de la scolarité sont mises sous la loupe et sujettes à de forts enjeux évaluatifs; la 8^e année (6P) présente, dans pratiquement tous les cantons romands, des évaluations de type sommatif, essentiellement utilisées pour l'orientation des élèves au secondaire; cependant, le poids qui leur est attribué pour la promotion de l'élève varie sensiblement d'un canton à l'autre (voir point 1.3.2).
- Certaines années sont vierges d'évaluation de type sommatif, laissant supposer qu'elles constituent un temps nécessaire et réservé au renforcement et à la stabilisation des apprentissages (3^e, 5^e et 7^e années).

1.3.1 Évaluations cantonales au cycle 1

En analysant plus finement les épreuves cantonales prévues au cycle 1 (voir annexe 8.2), on peut constater que :

- Le cycle 1 comporte très peu d'épreuves. Cela semble démontrer une volonté partagée de laisser du temps aux apprentissages essentiels de ces premières années de scolarité.
- Les quelques épreuves, principalement à visée diagnostique, sont situées plutôt en fin de cycle et semblent avoir majoritairement pour intention la mise en évidence de certaines difficultés d'élèves qu'il s'agira de surmonter dès l'entrée au cycle suivant.

1.3.2 Évaluations cantonales au cycle 2

En analysant plus finement les épreuves cantonales prévues au cycle 2 (voir annexe 8.2), on peut constater que :

- La fin de la 6^e année (4P) semble être un jalon pour le cycle ; en effet, tous les cantons à l'exception de Berne francophone possèdent, à cet endroit, des épreuves à visée soit diagnostique, soit sommative. Il semble que certains apprentissages doivent pouvoir être assurés pour la fin de cette année scolaire, sans lesquels la suite du cursus pourrait être sérieusement compromise.
- La 8^e année (6P) est clairement l'année d'orientation pour tous les cantons, certaines procédures mises en place étant plus ou moins lourdes et tenant plus ou moins compte des résultats aux épreuves. Il est également à relever que les différences entre ces diverses procédures d'orientation dépendent fortement d'organisations et de cultures cantonales (comme par exemple la prise en compte différente des résultats d'un élève selon l'établissement où il s'inscrit pour entrer à l'école secondaire).

1.3.3 Évaluations cantonales au cycle 3

En analysant plus finement les épreuves cantonales prévues au cycle 3 (voir annexe 8.2), on peut constater que :

- Les disciplines français, mathématiques et allemand, qui sont généralement les disciplines de promotion, sont le plus souvent évaluées durant ce cycle. À part le canton de Berne, tous les autres cantons proposent des épreuves en français et en mathématiques au moins une fois durant le cycle.
- Ce cycle présente une grande variété de situations ; il existe autant de cantons qui n'ont aucune épreuve (ou éventuellement de type diagnostique) que de cantons qui en font passer chaque année.
- Dans certains cantons, l'évaluation en fin de 11^e a une visée certificative alors que dans d'autres, le résultat n'est utilisé que dans une moindre mesure pour la promotion de fin d'année.

Au-delà de la diversité des situations cantonales, il ressort une spécificité de l'évaluation pour chaque cycle : une prédominance du diagnostique au cycle 1, de l'orientation au cycle 2, une diversité et un grand nombre de pratiques au cycle 3.

Relevons encore que lorsque les épreuves sont inscrites dans les lois cantonales, elles le sont diversement. Par exemple, le canton de Genève inscrit explicitement chacune de ses épreuves dans la loi cantonale, alors que Fribourg en inscrit simplement le principe.

1.4 Interprétations de l'article 15 de la Convention scolaire romande (CSR)

Dans la CSR, l'évocation d'épreuves romandes, que ce soit dans l'introduction, dans l'article 15 ou dans ses commentaires appellent des clarifications. Les échanges à leur sujet, dans le *Consortium romand* comme dans les *Groupes de travail*, montrent rapidement la nécessité de construire un langage évaluatif commun (voir point 2.3). Ainsi,

la distinction des définitions d'évaluation à visée sommative et d'évaluation à visée diagnostique (voir point 1.3) s'est-elle avérée utile pour comprendre la marge d'interprétation possible de l'article et de ses commentaires.

Article 15 – Épreuves romandes

¹ La CIIP organise des épreuves romandes communes à l'espace romand de la formation en vue de vérifier l'atteinte des objectifs du plan d'études.

² En fin de cycle ou à la fin du secondaire I, si la discipline choisie pour l'épreuve romande commune correspond à celle d'un test de référence vérifiant un standard national, le test de référence peut servir d'épreuve commune.

L'alinéa 1 de l'article sur les épreuves romandes fait apparaître la volonté politique de vérifier l'atteinte des objectifs du plan d'études.

Or, mesurer la qualité d'un système passe généralement par des épreuves bilans (voir point 2.3). Toutefois, le terme « épreuve commune » prête à confusion car il laisse sous-entendre qu'il s'agit d'une même épreuve présentée à tous les élèves, ce qui n'est ordinairement pas le cas pour des épreuves bilan; celles-ci, en effet, utilisent des méthodologies statistiques appliquées à des épreuves qui ne sont pas les mêmes pour tous les élèves, la somme des questions permettant alors de couvrir l'ensemble du référentiel que l'on souhaite vérifier. C'est le cas pour PISA et les tests de référence HarmoS. Ces enquêtes ne permettent toutefois pas de comparer des résultats individuels ni de fournir des informations pour chaque élève.

L'alinéa 2 établit un lien entre les tests de référence (HarmoS) et les épreuves romandes.

Utiliser le test de référence en lieu et place d'une épreuve romande impliquerait une parfaite adéquation entre eux. Selon la constitution actuelle des tests de référence, cela signifierait que les épreuves seraient avant tout destinées au monitoring, donc passées à des échantillons d'élèves (comme cela est prévu dans le cadre d'HarmoS).

Dans les cantons, bien que certaines épreuves soient à visée sommative (par exemple une épreuve d'orientation) donc prévues pour informer l'élève et l'enseignant en priorité, les résultats sont toutefois rassemblés et agrégés pour fournir une information globale utile au système. Si cette façon de faire est envisageable, elle exige néanmoins des précautions au moment d'interpréter les résultats et d'en extraire des généralités.

L'utilisation des résultats d'une première épreuve à d'autres fins que celles pour lesquelles elle est prévue n'est possible que dans la mesure où les informations fournies sont plus vastes que celles qui seraient obtenues dans une seconde épreuve. En d'autres termes, qui peut le plus peut le moins, avec pour conséquence qu'il ne saurait être question d'utiliser les résultats d'épreuves bilan pour en extraire une information similaire pour chaque élève; en effet, tous les élèves ne sont pas soumis au même cahier de test, ce dernier ne couvrant qu'une partie des objectifs visés.

Le sujet du présent rapport n'est pas de déterminer quelles seront les visées d'épreuves romandes et la manière dont seront traités les résultats. Nous concluons donc cette partie sur les options à prendre politiquement, actuellement en cours de traitement dans la constitution du règlement d'application de la CSR, par d'ultimes remarques et,

ci-après, par la mise à plat de deux scénarios plausibles et retenus par le *Consortium romand*:

- la réflexion engagée au niveau romand doit mener à proposer aux cantons des épreuves qui leur permettent de se situer relativement à l'atteinte des objectifs du PER;
- le *ranking* des établissements (liste nominative des établissements présentée dans un classement) et des enseignants sera impossible au niveau romand puisqu'on n'y disposera que de données cantonales anonymisées;
- le développement d'une banque d'items romande constitue une façon pertinente de mutualiser les expériences romandes en les ajustant au PER et, qui plus est, de contribuer à l'élaboration d'une banque d'items nationale.

1.5 Scénarios explorés

Afin d'explorer les différentes possibilités évoquées par l'article, le *Consortium romand* retient, dans un premier temps, les options suivantes:

- les épreuves romandes ne devraient pas faire doublon avec les tests de référence HarmoS;
- tous les élèves devraient être concernés par les épreuves romandes, *a priori* pour chaque cycle.

Se calquant sur les situations d'évaluation cantonales observées (voir point 1.3), deux scénarios sont alors explorés: l'un propose des épreuves romandes à visée sommative et l'autre des épreuves romandes à visée diagnostique. Ces scénarios tiennent compte des contraintes opérationnelles qui leur sont propres mais n'indiquent pas comment utiliser les résultats obtenus: cette décision appartient à chaque canton et il ne s'agit pas ici de viser une uniformisation des pratiques.

1.5.1 Scénario 1 : évaluation sommative

Le premier scénario explore la possibilité de disposer, en fin de chaque cycle, d'épreuves romandes à visée sommative, que leurs résultats soient destinés à être utilisés – peu ou prou – dans la promotion de l'élève ou dans l'orientation, selon le choix des cantons.

Qu'elles soient utilisées à des fins d'orientation ou de promotion des élèves, ces épreuves portent sur un même ensemble d'objectifs évalués et sont réalisées selon des modalités de passation semblables; les améliorations qu'elles peuvent permettre ensuite concernent essentiellement les pratiques enseignantes mais rarement et directement les apprentissages des élèves.

Développer des épreuves sommatives en fin de cycle nécessite de remplir un certain nombre de conditions: d'une part la passation doit avoir lieu en avril ou en mai; d'autre part, il s'agit de choisir consensuellement parmi les objectifs du PER ceux à évaluer dans chacune des disciplines concernées et de développer des questions dont le niveau de difficulté est éprouvé (adapté à ce qu'on attend de tous les élèves en fin de cycle).

De toutes, la plus forte des contraintes est certainement celle qui consiste à assurer une passation simultanée dans toutes les classes, sur la base de conditions de passation identiques pour tous.

Envisager des épreuves romandes à visée sommative pour tous les élèves en fin de chaque cycle présente plusieurs avantages :

- faciliter, du point de vue politique, la communication au niveau romand concernant la nature et les intentions de ces épreuves ;
- certifier l'atteinte, par les élèves, d'objectifs représentatifs du PER et d'en informer les parents, renforçant, *a priori*, la compréhension et l'adhésion à un référentiel romand, par extension à l'espace romand de la formation ;
- donner une information externe aux enseignants, transmissible au cycle suivant, ayant également un effet sur l'harmonisation de l'interprétation du PER ; cet effet sera manifeste à condition d'assurer un accompagnement dans l'interprétation de ces résultats et de rendre visibles les améliorations à apporter dans les pratiques d'enseignement ;
- permettre aux enseignants de comparer les résultats de leur classe à ceux d'un groupe plus vaste (par exemple le canton) ;
- informer le canton de l'atteinte globale, par les élèves, des objectifs testés ; bien que la mesure ne puisse fournir, chaque année, de véritables outils pour le monitoring, le suivi sur plusieurs années peut permettre d'évaluer l'évolution de l'harmonisation de l'enseignement en référence au PER ;
- accompagner le remplacement de certaines épreuves cantonales existantes ou de fournir une évaluation complémentaire à celles-ci, en fonction des besoins de chaque canton.

Bien que les avantages soient nombreux, les limites de telles épreuves sont non négligeables, avant tout parce que leur forte visibilité aux yeux du public nécessite de rendre ces épreuves irréprochables ; elles comportent les risques suivants :

- l'augmentation, pour certains cantons, du nombre de passations d'épreuves dans des années scolaires souvent déjà chargées d'enjeux de promotion ;
- l'ampleur de la coordination à assurer dans la mise en œuvre d'une épreuve romande et la lourdeur de passations annuelles (fréquence nécessaire) ;
- l'aspect critiquable de n'évaluer que globalement certains objectifs, ne rendant compte que partiellement des capacités réelles des élèves en fin de cycle ;
- des utilisations inadéquates ou variables des résultats par rapport à la fonction initialement attribuée, notamment une comparaison des établissements, et
- la difficulté de transmettre l'information aux enseignants de l'année suivante.

En conclusion, si le scénario d'une évaluation romande sommative peut être séduisant par sa simplicité et la bonne visibilité qu'il procurerait à l'harmonisation romande, il nécessiterait des travaux importants et une coordination considérable entre les cantons. Ceux-ci auraient également à assurer, dans leur organisation interne, une adéquation et une cohérence avec les évaluations cantonales existantes.

1.5.2 Scénario 2 : évaluation diagnostique

Le second scénario explore la possibilité de créer des épreuves diagnostiques. Elles seraient également prévues pour tous les élèves, et leur fonction première serait de repérer, à l'issue d'un cycle, des difficultés relatives à des objectifs du PER existant encore pour certains d'entre eux.

Pour un temps de passation identique à celui accordé à des épreuves sommatives, de telles épreuves portent toutefois sur une quantité limitée d'objectifs. En effet, le but étant de repérer des difficultés chez l'élève, il s'agit d'obtenir des informations plus précises sur un même objet d'apprentissage, en restreignant ainsi le nombre des objets observés.

Les conditions à remplir pour ce genre d'épreuves sont moindres que pour des épreuves sommatives. En effet, si le moment de passation se situe de préférence durant le début de l'année qui suit directement la fin d'un cycle, il n'existe pas de nécessité que tous les élèves soient soumis au même cahier au même moment.

Par ailleurs, si les cantons ont à s'accorder sur des éléments du PER à évaluer au terme des cycles 1 et 2, il n'est pas fondamental que cet accord soit total, dans la mesure où l'évaluation diagnostique ne porte pas les mêmes enjeux de promotion que l'évaluation sommative.

Une solution différente resterait à trouver pour le cycle 3 qui marque la fin de la scolarité obligatoire.

Outre la souplesse organisationnelle de ce scénario, d'autres avantages méritent d'être relevés :

- ne pas surcharger les fins de cycle ;
- informer l'élève (et ses parents) de l'atteinte de certains objectifs dans le but de remédier, avec l'encadrement adéquat, à ses difficultés ; idéalement, ces remédiations devraient également permettre de développer les stratégies d'apprentissage en cohérence avec les capacités transversales du PER ;
- outiller les enseignants pour procéder à des remédiations auprès des élèves en difficulté voire leur proposer des pistes d'amélioration de leurs pratiques ;
- permettre aux enseignants de comparer les résultats de leur classe à ceux d'un groupe plus vaste (par exemple le canton) ;
- informer le système sur le niveau d'atteinte de certains éléments du PER, ou sur d'éventuelles difficultés récurrentes chez les élèves, en vue d'un réajustement éventuel de certains éléments du référentiel (suivi du PER) ;
- vérifier que les questions proposées dans les épreuves sont en adéquation avec le PER tel qu'il est compris et travaillé sur le terrain d'une part, et avec le niveau des élèves d'autre part ; ceci est d'autant plus réalisable qu'une question mal ajustée dans une épreuve diagnostique peut simplement être écartée sans que cela invalide l'épreuve (contrairement à une épreuve sommative où l'épreuve entière pourrait être remise en cause).

Sans qu'il soit question de monitoring à proprement parler, il est possible, à partir de résultats obtenus sur plusieurs années, de faire quelques constats sur la progression des pratiques et leur adéquation avec les contenus du PER.

Une des difficultés d'un tel dispositif est de faire comprendre à toutes sortes de publics ce qu'il est et d'en expliquer l'intérêt. Il présente en outre trois inconvénients importants :

- le risque d'une augmentation globale des épreuves sur toute la scolarité, dans la mesure où, pour les cantons les mieux dotés en épreuves cantonales, essentiellement sommatives, elles viendraient s'y ajouter sans pouvoir les remplacer ;
- la dérive de l'utilisation des résultats à des buts d'évaluation sommative ; ce genre de dérive existe d'autant plus que l'investissement consenti par les enseignants pour mener ce genre d'évaluation est important et pourrait alors les amener à vouloir lui donner une portée plus large que celle de l'évaluation diagnostique ;
- la mise en place d'un accompagnement formatif des enseignants, à l'instar de pratiques existant dans certains cantons.

Le développement d'épreuves diagnostiques nécessiterait, outre le développement des questions, un travail de grande envergure conduisant à la réalisation d'outils d'information et de remédiation destinés aux enseignants. Nous relevons que, dans la situation actuelle d'implantation du PER, le développement d'épreuves diagnostiques pourrait constituer un outil intéressant pour l'interprétation de ce référentiel d'enseignement en termes évaluatifs.

2. Réflexions générales

Cette partie présente, essentiellement à travers les préoccupations romandes actuelles, à la fois l'évolution que subissent les référentiels d'enseignement, désormais rédigés comme des curriculums, et les conséquences que cette évolution peut induire sur l'évaluation. Il nous semble important d'aborder ces deux points dans la mesure où le PER est le document de référence sur lequel nous bâtissons nos épreuves prototypiques (une en mathématiques, une en français), et où nous voulons développer un modèle d'évaluation pour la construction d'épreuves.

De nombreux ouvrages abordent ces problématiques et le sujet ici n'est pas d'en rendre compte de manière exhaustive mais d'extraire des différents modèles existants les éléments les plus pertinents dans le cadre de nos travaux de réflexion sur l'élaboration d'épreuves romandes communes. Nous aborderons également la question des fonctions de l'évaluation en nous appuyant notamment sur les travaux réalisés en Suisse romande.

2.1 L'évolution des référentiels d'enseignement

Avec l'accord des responsables politiques, les concepteurs du PER ont choisi de présenter le référentiel d'enseignement sous forme de curriculum, option déjà prise dans les plans d'études précédents de plusieurs cantons romands. Selon Frey (1986), *le curriculum, (...) est une conception de l'enseignement qui en fait, à l'intérieur d'une période donnée, un système cohérent, comportant des procédures multiples pour la planification, la réalisation optimale et le contrôle des résultats*. Plusieurs facteurs vont conduire les pays européens à adhérer à la notion du curriculum développée aux États-Unis. Parmi les principaux, citons l'avènement de la compétence dans le champ scolaire; la professionnalisation du métier d'enseignant; les échanges internationaux, la mondialisation et la globalisation conduisant à la migration des travaux scientifiques; et le développement de l'idéologie néolibérale promouvant une vision pragmatique et instrumentale du curriculum (Lenoir, 2011).

Dans cette optique, le curriculum doit en particulier permettre l'insertion de l'individu dans la société tant par des valeurs que des savoir-faire socialement reconnus. La situation d'apprentissage y est considérée comme un élément central dans lequel s'ancrent les savoirs vus surtout comme des outils, des instruments au service des compétences. Dans ce sens, l'action est primordiale (comment faire pour résoudre le problème posé dans la situation?). C'est dans ce contexte que le PER a été élaboré. Cependant, il présente certaines particularités qui méritent d'être soulignées.

Le PER développe un projet global de formation, avec des accents disciplinaires importants (voir point 3.1), encourageant aussi une éducation culturelle et humaniste et non pas uniquement fonctionnelle et pragmatique comme le sont les curriculums étatsuniens (Lenoir, 2011). Le risque évident est un trop-plein qui pourrait occasionner une impossibilité de poursuivre les objectifs d'apprentissage visés.

Mettant en avant le développement de compétences comme une finalité essentielle de l'enseignement obligatoire, une des ambitions du PER est de réduire l'analphabétisme fonctionnel qui, comme Roegiers (2004, p. 8) le rappelle, *est le mécanisme par lequel des jeunes qui ont fréquenté l'école durant plusieurs années ne sont pas capables de réinvestir leurs acquis dans les situations de la vie quotidienne (...)*. Il s'agit donc de faire en sorte que les élèves dépassent le seul répertoire des connaissances qui étaient prioritaires dans les programmes précédents.

Nous nous positionnons en faux contre l'idée fréquente qu'un tel curriculum serait trop ambitieux pour certaines catégories d'élèves: de nombreux travaux (Rey, Carette, Defrance, Kahn, 2004; Roegiers, 2004, ...) montrent que l'approche par les compétences permet de redonner du sens aux apprentissages de connaissances (auxquels certains élèves s'opposaient précisément parce qu'ils ne le percevaient pas), et que la résolution de situations complexes, pour autant qu'elles soient bien choisies, profite à tous les élèves mais davantage encore aux élèves faibles (Allal, Rouiller, Saada-Robert & Wegmuller, 1999; Aden & Roegiers, 2003). Les didactiques disciplinaires actuelles nous semblent aller dans ce sens en favorisant les approches centrées sur des résolutions de problèmes ou des situations complexes **pour tous les élèves**. Certains travaux (Guignard & Hirsig, 1989; 1991) démontrent les avantages de cette approche tant pour les bons élèves que pour ceux en difficulté. Pour autant que les situations soient ajustées, que l'accompagnement soit adéquat et que le temps à disposition soit suffisant, les élèves en difficulté peuvent tout autant venir à bout de situations complexes, avec pour effet collatéral de constater leurs capacités et de renforcer leur propre estime.

Enfin, il faut rappeler que le PER est le résultat d'un travail au sein d'une vaste équipe de rédacteurs provenant des sept cantons de Suisse romande. Il a donc fait l'objet de multiples discussions, modifications, compromis. Finalement, il a donné lieu à une large consultation dans les cantons et a été adopté par la CIIP/SR-TI. Tous les partenaires de l'école ont été associés à la consultation et ont pu donner leur avis sur son organisation et son contenu. Le PER témoigne ainsi d'une adhésion sociale. Il reflète certaines sensibilités cantonales ou propres à tel partenaire consulté. Dans ce sens, il n'est pas un document idéologique élaboré par une équipe d'experts, mais davantage le résultat d'une construction collective incluant les différents acteurs de l'école (enseignants, formateurs, chercheurs, didacticiens, parents, responsables scolaires,...).

Les compétences

Nous avons souligné que l'une des spécificités des curriculums actuels, dont le PER, est le développement de compétences chez tous les élèves, tout au long de la scolarité obligatoire. Il nous semble dès lors important de comprendre ce qu'il faut entendre par compétence. De nombreux travaux traitent de cette notion et le nombre des définitions ne cesse de croître avec le temps. Peu à peu, elles incluent certains aspects dont on peut discuter la pertinence sur un plan scolaire: d'ordre motivationnel, ou volitif, par exemple (Weinert⁵, 2001). Si de tels éléments peuvent s'intégrer aux apprentissages, leur évaluation pose en revanche des problèmes cruciaux, en particulier quant à l'utilisation qui pourrait être faite des résultats obtenus.

⁵ *Les capacités et aptitudes cognitives dont l'individu dispose ou qu'il peut acquérir pour résoudre des problèmes précis, ainsi que les dispositions motivationnelles, volitives et sociales qui s'y rattachent pour utiliser avec succès et responsabilité les résolutions de problèmes dans des situations variables.*

Avant d'entrer plus en détail dans ces définitions, il faut savoir que la notion même de compétence fait l'objet d'âpres discussions au sein des chercheurs et soulève la controverse (Actes du colloque de l'ADMEE, 2012⁶). Sans entrer dans la polémique, il convient néanmoins de relever quelques-unes des principales critiques apportées à cette notion. Il apparaît souvent un certain flou autour de cette notion, difficile à distinguer de comportement, attitude, savoir-faire, connaissance, qualification ou capacité ; par ailleurs, très présente dans le monde professionnel, il n'est pas certain que la notion de compétence puisse d'emblée s'adapter au monde scolaire sans que se pose la question de sa définition, des conséquences sur l'enseignement-apprentissage, ses finalités, et sur l'évaluation des élèves (Bronckart & Dolz, 2000 ; Bain, 2000). Or, le rôle de l'école est bien de permettre l'acquisition de savoirs constitués qui se réfèrent non seulement aux pratiques sociales, mais également aux savoirs scientifiques ou d'experts. Autrement dit, ce rôle ne se cantonne pas à développer chez les élèves des comportements utiles, des savoir-faire, il doit les conduire à transformer leurs connaissances (Perrenoud, 1998, 2002). Les savoirs ne sont donc pas seulement au service de l'agir mais contribuent tout autant, par un travail de réflexion, à construire un rapport à la connaissance.

La notion de compétence présuppose en outre l'idée d'un transfert des apprentissages d'une situation donnée à d'autres situations. Or, il faut bien l'admettre, nous savons peu de choses sur ces phénomènes de transfert dans les apprentissages. De plus, il n'est pas certain que des connaissances acquises dans une situation soient celles qu'il s'agirait de mobiliser dans une autre situation, même proche. Cela rend le statut de la compétence fragile. Ce qui fait dire à Jonnaert (2003, p. 24) que (...) *l'approche du concept de compétence en éducation s'est aujourd'hui fondamentalement désenclavée de la définition de la compétence linguistique, de celle de la performance (...)*. Si dans les années 1950 Chomsky a redéfini la notion de compétence dans le domaine linguistique de façon très spécifique et a avec ses travaux fortement influencé la linguistique, ainsi que d'autres sciences humaines et sociales, il faut bien prendre en compte que la notion de compétence en sciences de l'éducation est antérieure aux travaux de Chomsky. Allal (2002) note que cette notion *est présente depuis plus d'un siècle dans les travaux en psychologie, en psychométrie, en linguistique et qu'elle a influencé dès le départ les cadres conceptuels des sciences de l'éducation* (p. 77).

Il n'en demeure pas moins que, puisque selon le mandat qui nous est donné nous avons à nous référer au PER, il nous faut nous emparer de cette notion. Dans une approche empirique qui est celle adoptée dans le cadre de notre travail, nous avons fait le choix de partir de la définition fournie par le PER :

« Possibilité, pour un individu, de mobiliser un ensemble intégré de ressources en vue d'exercer efficacement une activité considérée généralement comme complexe ». (PER, 2010)

Nous constatons que cette définition est très proche de celles de nombreux chercheurs, comme par exemple Roegiers (2004) :

« Une compétence est la possibilité, pour un élève, de mobiliser un ensemble de savoirs, savoir-faire et de savoir-être pour résoudre des situations. » (p. 107)

⁶ 24^e colloque international de l'ADMEE - Europe. *L'évaluation des compétences en milieu scolaire et en milieu professionnel*. <http://admee2012.uni.lu/pdfdownload.html#5>

Cette définition, comme celle fournie par le PER, fait référence d'une part à la mobilisation de ressources – c'est-à-dire de savoirs, savoir-faire et savoir-être – qui, dans notre cas, sont précisés dans le PER, d'autre part à la résolution de situations ou d'activités complexes.

Quelques précisions complémentaires sont données dans la définition suivante :

« Une authentique compétence est la capacité à répondre à des situations complexes et inédites par une combinaison nouvelle de procédures connues; et non pas seulement à répondre par une procédure stéréotypée à un signal préétabli. » (Rey & al., 2006, p. 26)

Nous retenons de la définition de Rey le fait que les situations complexes sont inédites et que les ressources (identifiées à « procédures connues ») nécessitent d'être choisies et combinées. Selon Crahay (2006), le côté inédit de la situation complexe présente un danger pour les élèves dans la mesure où elle les confronte à un haut niveau d'exigence. Le fait même de devoir se faire une représentation de la situation nouvelle, à quoi s'ajoutent le choix et la mobilisation des procédures de résolution du problème, constituent, pour lui, une source de surcharge cognitive à même de placer les élèves en échec. Il s'agira donc pour nous, dans le cadre évaluatif qui est le nôtre, de nous poser la question du choix de la situation à proposer aux élèves et de son degré de nouveauté pour eux. En ce sens, tout l'enjeu réside dans le choix ajusté d'une situation d'évaluation qui se doit d'être (...) *juste assez inédite pour que la pure et simple répétition soit inadéquate, juste assez familière pour que le sujet ne se sente pas totalement démuné*, comme Perrenoud⁷ l'écrit (Crahay 2006, p. 103). Nous reviendrons sur ces aspects dans la partie 4.1 lorsque se présentera la concrétisation d'une épreuve.

Par ailleurs, il s'agit de souligner que les « situations complexes » définies par Rey présentent entre autres une différence particulière par rapport aux deux premières définitions que nous avons citées : Rey et al. (2004) proposent en effet des situations pluridisciplinaires qui, si elles ne permettent pas aux élèves d'identifier la tâche à réaliser dans une discipline en particulier, s'avèrent idoines, selon les auteurs, pour faire émerger les compétences que l'on veut évaluer. Bien que les raisons invoquées soient tout à fait pertinentes, nous prenons quant à nous l'option d'identifier des situations complexes disciplinaires. En effet, des compétences spécifiques sont déclinées dans les disciplines présentes à l'intérieur des domaines du PER. Il nous semble dès lors important de choisir, pour l'évaluation, des situations à même de fournir des informations sur ces compétences et sur les connaissances et savoir-faire disciplinaires mobilisés tout en réduisant, dans la mesure du possible ou si nécessaire, l'impact d'éléments provenant d'autres disciplines.

Afin d'illustrer notre propos, notons par exemple que l'utilisation de la langue de communication, le français en l'occurrence, est présente dans toutes les situations. Ainsi, la mobilisation de ressources en français nécessaires à la résolution d'un problème, par exemple en mathématiques, ne devrait pas être une entrave pour l'élève et nécessite d'en réduire l'impact, notamment en ne faisant appel qu'à des ressources en français qui seraient considérées comme maîtrisées. Ainsi nous paraît-il envisageable de

⁷ Bien qu'il ne mentionne pas clairement le cadre d'utilisation de ce genre de situation (apprentissage versus évaluation), ce que déplore Crahay, la formulation que nous lui empruntons nous semble pertinente pour décrire une situation d'évaluation et permet de rendre visible toute la difficulté du choix d'une telle situation.

produire de l'information relativement à un objet d'évaluation tel que *Résoudre des problèmes additifs* figurant dans le PER.

Ce choix nous semble d'autant plus pertinent que les élèves sont jeunes ; il pourrait être repensé différemment lors de l'élaboration d'épreuves pour la fin de la scolarité. Toutefois, il paraît délicat, dans la création d'une épreuve telle qu'il nous est donné de la concevoir, d'annoncer que nous évaluons la compétence de l'élève à mobiliser ses connaissances, toutes disciplines confondues, et ce à plusieurs égards :

- être compétent peut-il vraiment être considéré comme une compétence alors que les clés de son développement ne sont pas données aux enseignants, qui plus est en fin de scolarité au moment où les enseignements se font souvent par des spécialistes ?
- les disciplines, encore principales organisatrices de la scolarité, possèdent des compétences déjà largement complexes qu'il s'agit d'explorer ; par ailleurs, à travers les épreuves élaborées, il serait utile d'y associer, dans la mesure du possible, des résultats et pistes de remédiation ;
- la fin de la scolarité obligatoire n'est plus, comme ce fut longtemps le cas, la fin des apprentissages ; les formations subséquentes prolongent amplement les apprentissages, parfois dans des domaines pluridisciplinaires, mais également dans les mêmes disciplines que celles du secondaire 1.

Nous retiendrons ainsi les définitions du PER, de Roegiers et de Rey en précisant les choix suivants relatifs à notre projet :

- Nous considérons les ressources comme l'ensemble des savoirs, savoir-faire et savoir-être scolaires – entendu par là que ces derniers sont travaillés en contexte scolaire et bénéficient donc d'un apprentissage spécifique disciplinaire – qui sont à mobiliser en situation ; le niveau de maîtrise de ces ressources est à considérer relativement au référentiel, bien qu'une marge d'interprétation demeure, le PER étant un référentiel d'enseignement et non d'évaluation.
- Nous retenons le principe que l'expression d'une compétence se fait à travers des situations complexes, en rajoutant une dimension disciplinaire à ces situations.

À noter que nous faisons le choix de réserver le terme procédure à la démarche de résolution d'un problème ou la réalisation d'une tâche complexe par l'élève. Cela nous permet de caractériser les réponses des élèves en termes de graduation des procédures engagées, expertes, semi-expertes ou exploratoires.

2.2 L'évaluation de compétences

Le travail principal de ce projet est avant tout de caractériser l'évaluation la plus à même de fournir des informations sur certains éléments, essentiellement disciplinaires, du curriculum qu'est le Plan d'études romand. L'analyse de ces éléments montre qu'ils représentent autant des savoirs et savoir-faire que des compétences. Au-delà de la réflexion sur l'articulation entre ces éléments à prendre en compte dans une évaluation (voir chapitre 3), se posent au préalable d'une part la question des degrés et niveaux de la compétence et d'autre part celle de sa mesure.

Il est possible de trouver dans la littérature traitant des compétences des exemples de formulations de compétence d'un niveau proche de celles du monde professionnel dont elles sont originaires. Par exemple, *faire un rapport* est une formulation qu'il est possible de rencontrer dans nombre de professions et nombre de référentiels d'enseignement; *faire un rapport* peut être considéré effectivement comme une compétence, très vaste certes, mais dont la formulation est oh combien pratique par sa simplicité, et universelle par le nombre de contextes dans lesquels elle peut s'exprimer. Or, nous pensons qu'il s'agit là avant tout d'une formulation regroupant des compétences communes, sorte de famille de compétences, et qu'elle est clairement non atteignable et encore moins maîtrisable par un seul individu dans tous les domaines. En règle générale, le contexte dans lequel cette compétence est appelée à s'exprimer, qu'il soit professionnel ou scolaire et disciplinaire, déterminera, dans l'action, des réflexions et productions fort différentes: on ne fait pas un rapport en physique comme l'on fait un rapport de recherche en biologie par exemple. Qui plus est, l'interlocuteur à qui sera destiné le rapport influencera largement sa forme, son niveau de précision, sa structure, le vocabulaire employé, les symboles, les références, etc. Il est certain que l'utilisation du mot « rapport » influence largement l'idée d'une parenté de compétences, mais il est également vrai que des traits communs peuvent être relevés entre elles. Dans chacune des situations où l'on peut dire qu'un « rapport est fait », on constatera la nécessité d'une analyse, d'un apport théorique permettant de raccrocher la situation à des faits similaires et d'une rédaction d'un texte dans l'optique d'un « rendre compte ».

De fait, l'esprit humain tend à trouver des règles générales dans l'organisation du monde l'entourant, qu'il soit issu de la nature ou de son propre esprit. Si on peut décemment penser que des formulations aussi générales que « faire un rapport » sont plus qu'aisées pour communiquer, elles ne sont toutefois pas suffisantes pour déterminer les actions et réflexions qu'un individu devra mener pour faire acte de cette compétence, qui se fera dans un contexte bien déterminé et non dans l'absolu.

Ainsi, comme le suggère Roegiers (2004), la situation, ou mieux encore, les situations déterminent-elles la compétence; nous affirmons même qu'il est impossible de recueillir toutes les situations où cette compétence serait susceptible de s'exprimer et qui permettraient alors vraiment de certifier qu'un individu « possède » cette compétence. Le champ des possibles situations où peut s'exprimer une compétence étant si vaste, on ne peut que déclarer qu'un individu, placé dans l'une d'elles, sera amené à « développer » cette compétence⁸. Le contact répété à ce genre de situations devrait le conduire à un « niveau » de développement de cette compétence que l'on pourrait alors estimer suffisant et qui sera susceptible d'augmenter lors de confrontation à toute nouvelle situation de ce genre. Vouloir attester d'une compétence ne relèverait-il pas d'un fantasme d'une mesure exacte des possibilités d'un individu ?

Dans ces conditions, nous faisons l'hypothèse qu'une évaluation de compétences n'apporte que des informations partielles sur les compétences visées et ne peut donc attester d'une atteinte complète et définitive de ces compétences.

⁸ Une situation pouvant également nécessiter la mise en oeuvre de plusieurs compétences.

Des degrés de compétence et des niveaux de compétence

Venons-en à la première question que nous soulevions plus haut, relative aux degrés et aux niveaux de compétence. Rey et al. (2004), proposent trois degrés de compétence :

- « *compétence de premier degré* : savoir exécuter une action en réponse à un signal préétabli, après entraînement ; appellation : *compétence élémentaire ou procédure* ;
- *compétence de deuxième degré* : savoir choisir, parmi les procédures qu'on connaît, celle qui convient à une situation ou à une tâche non connue ; appellation : *compétence élémentaire avec interprétation (ou cadrage) de la situation* ;
- *compétence de troisième degré* : savoir, parmi les procédures qu'on connaît, choisir et combiner celles qui conviennent à une situation ou à une tâche non connue et complexe ; appellation : *compétence complexe*. »

À l'instar de ces auteurs, nous assimilons les *compétences du troisième degré* à ce que nous définissons comme une compétence et les *compétences du premier degré* à des ressources. Les *compétences du deuxième degré*, bien qu'elles tentent de caractériser un entre deux, semblent difficiles à exploiter telles quelles. Les résultats de la seconde phase d'évaluation réalisée par l'équipe belge portant précisément sur des compétences du deuxième degré, semblent d'ailleurs corroborer notre doute (voir point 3.2). Toutefois, la graduation proposée de la compétence fournit une clé d'analyse pour interroger les éléments présentés dans le PER et nous permettent de mieux les appréhender.

Précisons que la formulation d'une compétence ne détermine pas univoquement un degré à cette compétence (Rey & al., 2006). Celui-ci dépend en effet de l'âge de l'élève et de l'état de ses connaissances. Dans le développement de compétences réside également l'idée d'intégration de ces compétences à travers le temps, faisant passer des compétences de degré 3 à des compétences de degré 2 voire 1. C'est ce qui permet à l'élève de progresser et d'accéder à des compétences toujours plus complexes, sans toutefois prétendre qu'une compétence de degré 2 ou 1 soit forcément parfaitement atteinte ; il peut demeurer comme le mentionne Roegiers (2004), des cas où, malgré un constat maintes fois renouvelé d'une compétence atteinte, un élève face à une nouvelle situation ne parvienne pas à mobiliser cette compétence. Par exemple, *accorder les pluriels*, qui est en début de scolarité une compétence plutôt de degré 3, tend-elle à devenir une compétence de degré 2 voire idéalement 1, sachant qu'il arrive encore à bien des scripteurs, même chez des adultes expérimentés, d'omettre certains accords de pluriels dans un texte.

Cette première analyse peut également s'enrichir de la notion de niveaux de compétence. Bien que maintes expériences aient montré toute la difficulté de caractériser précisément des niveaux de compétences (PISA, HarmoS, PEL, ...), tenter de le faire nous semble tout de même intéressant dans la mesure où, comme expliqué plus haut, nous estimons impossible d'attester de l'atteinte complète d'une compétence. La définition de niveaux d'atteinte intermédiaires pourrait alors apporter des informations utiles. La majorité des tentatives dans ce sens se heurtent toutefois à la difficulté de définir des niveaux de compétence autrement que par l'utilisation d'un vocabulaire de graduation allant du simple au compliqué (ou complexe), du familier au général,

du court au long, et sans qu'il soit possible de donner un sens clair à ces adjectifs de graduation de façon à pouvoir comprendre ce qu'ils recouvrent.

Dans les recherches citées ci-dessus, la définition des niveaux de compétence sert essentiellement à positionner ensuite les élèves dans ces différents niveaux. Notre démarche est différente : nous cherchons à recueillir des informations sur les acquis et les difficultés des élèves en vue d'aider les enseignants à organiser leur enseignement. Par ailleurs, la déclinaison des compétences en sous-compétences peut conduire à la fois à une pléthore de compétences alors même qu'il est impossible d'affirmer que la somme des sous-compétences constitue la compétence première. On retrouverait ainsi un des inconvénients majeurs de la pédagogie par objectifs qu'est le morcellement excessif des capacités.

Des situations complexes

Nous avons vu plus haut qu'il n'y a pas d'évaluation de compétence sans tâches et situations complexes voire inédites. Selon Jonnaert (2003), *l'élément le plus important de cette architecture [des compétences] est (...) la situation. Elle est à la fois le point de départ de la compétence et le principal critère qui servira à vérifier si la compétence fut efficace ou non dans sa mise en œuvre* (p. 61). Nous avons vu aussi combien ces caractères de nouveauté et d'inédit sont problématiques pour certains ; Crahay (2006) craint en particulier une confusion possible entre situations d'apprentissage et situations d'évaluation. S'il peut sembler pertinent de proposer des situations inédites dans le cadre des apprentissages, l'accompagnement que ce cadre propose n'existe pas lors d'une évaluation. En effet, le côté inédit et nouveau de la situation d'évaluation de compétences demande une appropriation de l'élève. Il est alors à la fois évalué et en état d'apprentissage. Enfin, la situation nouvelle sous-entend un transfert de compétences d'une situation à une autre. Mais que sait-on au juste de ce transfert ? Les situations sont-elles suffisamment proches pour que de mêmes compétences s'y manifestent ? Sinon, qu'est-ce qui est transféré ? Peut-on être certain qu'il suffirait de confronter les élèves à toute une série de situations complexes et inédites pour qu'ils acquièrent les compétences visées ? Le fait d'être au clair sur les ressources mobilisées dans chacune des situations aide-t-il à construire cette certitude ?

Certains préconisent de constituer des familles de situations pour lesquelles les mêmes compétences sont à exercer. Ainsi Roegiers (2004) écrit-il : *À chaque compétence est associée une famille de situations-problèmes. C'est un ensemble de situations « cibles » dont chacune est une occasion d'exercer la compétence : une occasion d'un niveau de complexité suffisant (en conditions réelles) mais d'un niveau qui ne dépasse pas ce qui est attendu. Toutes ces situations sont dites équivalentes, c'est-à-dire interchangeables en termes de niveau de difficulté et de complexité* (p. 110). Reste à déterminer ces niveaux de difficulté et de complexité, ainsi que, dans la mesure du possible, les caractéristiques communes de ces tâches qui en font une famille !

À propos de la situation-cible, il est intéressant de remarquer que Roegiers la distingue de situations problèmes didactiques qui, elles, ont pour but de permettre aux élèves l'acquisition de nouveaux savoirs et savoir-faire en classe, tout en développant certaines compétences ; la situation-cible, quant à elle, *est le reflet d'une compétence à installer chez l'élève. Elle peut être considérée comme une occasion d'exercer la compétence chez l'élève, ou comme occasion d'évaluer s'il est compétent (pour cette*

compétence précise) (p. 109). Quelles sont les caractéristiques des situations-cibles ? En quoi diffèrent-elles des situations didactiques ? En choisissant et déterminant nos propres situations pour l'évaluation, nous serons conduits à définir ces caractéristiques (voir chapitre 3).

Pour éviter que les situations choisies pour l'évaluation des acquis des élèves ne recouvrent que partiellement l'ensemble des compétences visées, Roegiers suggère de commencer par définir des compétences avant de définir des situations et de subordonner les situations aux compétences ainsi définies. Cette procédure a été la nôtre comme nous le verrons plus loin. Il affirme encore le bien-fondé de créer des épreuves dans le cadre des disciplines, la situation interdisciplinaire ne parvenant pas, selon lui, à couvrir toutes les compétences visées. Là encore, notre option a été d'envisager nos épreuves d'évaluation à l'intérieur de disciplines (en l'occurrence français et mathématiques).

2.3 Les buts et visées de l'évaluation

Depuis les années 90, des rencontres romandes autour du thème de l'évaluation se succèdent, notamment (...) *pour faire progresser les approches et pratiques de l'évaluation formative dans les divers degrés de la scolarité obligatoire* (...) (GRETEL, 2004, p. 17). En 2003, la CIIP décide de traiter au niveau romand la question de l'évaluation du travail de l'élève, tâche confiée à un groupe temporaire intercantonal, le *Groupe romand pour l'évaluation du travail des élèves* (GRETEL). Son mandat est de *proposer à la CIIP un dispositif-cadre d'évaluation praticable dans les classes, compréhensible pour les parents et utile à la gestion des parcours scolaires de l'élève*. Le rapport qu'il livre en 2004 propose des lignes directrices sous forme de propositions ainsi qu'une série de définitions relatives aux fonctions de l'évaluation. Le consortium et les groupes de notre projet prennent, dans un premier temps, appui sur ces définitions qui présentent l'avantage d'être le résultat d'un premier consensus romand.

Les buts de l'évaluation sont nombreux, comme le rappelle Cardinet (1996) :

- Améliorer les décisions relatives à l'apprentissage de chaque élève.
- Informer sur sa progression l'enfant et ses parents.
- Décerner les certificats nécessaires à l'élève et à la société.
- Améliorer la qualité de l'enseignement en général.

Selon le but principal d'une évaluation, nous définissons donc des évaluations à visées distinctes :

Évaluation à visée sommative : 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 certificative : évaluation sommative qui conduit à une attestation officielle de compétences globales au terme d'une période de formation et peut contribuer à la promotion de l'élève.

Évaluation de sélection ou d'orientation : évaluation certificative permettant de pronostiquer, des niveaux de compétences chez les élèves, certains dépassant les exi-

gences définies pour la période de formation qu'ils achèvent; le but est de répartir les élèves dans les filières ou les niveaux subséquents.

Évaluation à visée diagnostique: elle 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.

Évaluation bilan: évaluation scientifique (parfois appelée évaluation système) qui doit permettre de recueillir, à des moments-clés de la scolarité, des données représentatives de la maîtrise de tous les objectifs du plan d'études et de fournir des éléments permettant de les analyser et de les comprendre de manière à pouvoir permettre le pilotage du système éducatif par les responsables scolaires. Par conséquent, l'évaluation bilan ne peut fournir d'information sur les élèves pris individuellement.

Relevons que selon Rey et al. (2004), toute évaluation devrait (...) *permettre d'établir un diagnostic des compétences ou des processus de l'élève (...)* (p.39) et ainsi être diagnostique; les auteurs reconnaissent toutefois la pauvreté des informations de cet ordre dans le cas d'évaluations sommatives ou certificatives. Pour notre part, nous préférons considérer que certaines évaluations – que Rey définirait comme formatives – sont construites précisément dans le but d'établir un diagnostic des difficultés des élèves. En ce sens, nous suivons le choix du GRETEL (2004) qui précise que (...) *Elles (les évaluations diagnostiques) visent à mettre à leur (les enseignants) disposition des outils d'évaluation et d'analyse, sous forme d'une banque d'outils documentés les mettant en mesure de repérer des difficultés, d'analyser leur nature et leur cause, d'ajuster leur enseignement et de les aider à faire des choix pédagogiques au regard de la forte hétérogénéité des classes* (p. 34).

Dans le cadre de ce projet, l'intérêt d'explorer le développement d'évaluations diagnostiques réside dans le retour sur apprentissages qu'elles devraient permettre et sur la démarche exploratoire qu'un tel développement suppose (voir point 4.4). Alors que la tendance est, dans de nombreux pays où se développent des standards, d'utiliser les tests qui leur sont associés comme éléments premiers de pilotage du système éducatif, il nous semble utile également de proposer aux enseignants des outils d'interprétation possible des résultats obtenus par leurs élèves. En effet, se focaliser sur les apprentissages et moins sur les résultats des tests semble un gage d'amélioration des performances selon Walkins (2010): *Children do better in their exams when their teachers focus on learning, rather than on test results*⁹.

En conclusion, les quelques éléments théoriques sur lesquels nous nous sommes arrêtés ici constituent une base sur laquelle ont été construits notre modèle évaluatif puis nos épreuves prototypiques. C'est notamment dans cette phase de concrétisation ainsi que lors de l'analyse de nos résultats que nous pourrions réinterroger ces différents apports théoriques. En effet, s'engager dans la caractérisation de situations d'évaluation, par exemple, montre vite ses limites lorsqu'elle n'est interrogée que sur le plan conceptuel; la mise à l'essai de notre modèle sur des compétences disciplinaires montrera la nécessité de l'affiner au regard des caractéristiques propres aux disciplines.

⁹ <http://www.ioe.ac.uk/43212.html>

3. Développement d'un modèle d'évaluation adapté au PER

Se pencher sur l'élaboration d'épreuves romandes dans le but de fournir des renseignements sur l'atteinte des objectifs proposés dans le PER (voir chapitre 1) nécessite d'abord de définir ce qu'il est pertinent de prendre en compte dans ce référentiel d'enseignement pour créer une évaluation. C'est dans ce but qu'au point 3.1 nous analysons la structure et les contenus du PER. Peut-on simplement transférer les éléments d'un référentiel d'enseignement pour en faire un référentiel d'évaluation ?

Comme évoqué dans le chapitre 2, le PER est un curriculum et se structure, pour chaque cycle et chaque domaine disciplinaire, à travers des compétences visées. Nous avons donc cherché des modèles évaluatifs tendant à rendre compte des compétences et, après analyse, avons choisi de nous arrêter plus longuement sur celui développé par Rey, Carette, Defrance et Kahn en Communauté française de Belgique (voir point 3.2). Leurs expériences et conclusions relatives à ce modèle nous ont permis, en tenant compte des contraintes liées à notre projet, de développer un modèle en deux phases, présenté au point 3.3.

3.1 Un référentiel : le PER

Le Plan d'études romand est structuré de façon similaire quel que soit le domaine concerné. Ainsi, les éléments que sont les *visées du domaine*, les *objectifs d'apprentissage*, les *progressions des apprentissages* et les *attentes fondamentales* donnent à voir du plus large au plus précis ce qui est à travailler avec les élèves et dans quel sens le faire. La grande cohérence de ce référentiel porte à s'interroger sur deux aspects : quels sont les éléments qu'il s'agit de prendre en compte pour créer une évaluation ? et peut-on alors créer un modèle évaluatif similaire pour chaque discipline ?



Les *Groupes de travail* et l'*Équipe de projet* se sont penchés sur ces deux questions et ont, travaillé parallèlement et conjointement sur un concept d'évaluation pour le français et pour les mathématiques. Ce travail conjoint, ainsi que l'évaluation prototype (voir chapitre 4) développée dans le cadre de ce projet, indiquent que le rapprochement dans les concepts évaluatifs est réalisable tout en mettant à jour les possibles et les limites.

Fig. 3.1.1: *Projet global de formation de l'élève selon le PER*

Le Plan d'études romand présente un projet de formation global de l'élève (voir figure 3.1.1.) qui se structure selon trois entrées principales :

- des *Domaines disciplinaires*,
- la *Formation générale* et
- des *Capacités transversales*.

La structure des domaines disciplinaires

Chacun des domaines disciplinaires présente une structure en cascade similaire, à savoir :

- Des **visées du domaine** qui synthétisent les finalités et orientations prioritaires des disciplines constitutives du domaine ; par exemple, pour le domaine Mathématiques et sciences de la nature (MSN) :

« Se représenter, problématiser et modéliser des situations et résoudre des problèmes en construisant et en mobilisant des notions, des concepts, des démarches et des raisonnements propres aux mathématiques et aux sciences de la nature dans les champs des phénomènes naturels et techniques, du vivant et de l'environnement, ainsi que des nombres et de l'espace. »

- Un réseau d'**objectifs d'apprentissage** (voir figure 3.1.2) qui présente, par cycle et par thématique, des compétences relativement larges à développer chez les élèves ; ces objectifs traduisent ainsi les visées du domaine et organisent les disciplines au sein du domaine. Le réseau ci-dessous, par exemple, concerne les mathématiques dans le domaine MSN et présente cinq objectifs par cycle :

RÉSEAU DES OBJECTIFS D'APPRENTISSAGE

	Espace	Nombres	Opérations	Grandeurs et mesures	Modélisation
Premier cycle	MSN 11 Expliquer l'espace... Mathématiques	MSN 12 Pouvoir et résoudre des problèmes pour construire et structurer des représentations des nombres naturels... Mathématiques	MSN 13 Résoudre une problématique arithmétique... Mathématiques	MSN 14 Comparer et ordonner des grandeurs... Mathématiques	MSN 15 Représenter des phénomènes naturels, techniques ou des situations mathématiques... Mathématiques / Sciences de la nature
Deuxième cycle	MSN 21 Pouvoir et résoudre des problèmes pour structurer le plan et l'espace... Mathématiques	MSN 22 Pouvoir et résoudre des problèmes pour construire et structurer des représentations des nombres naturels... Mathématiques	MSN 23 Résoudre des problèmes arithmétiques et multiplicatifs... Mathématiques	MSN 24 Utiliser la mesure pour comparer des grandeurs... Mathématiques	MSN 25 Représenter des phénomènes naturels, techniques, sociaux ou des situations mathématiques... Mathématiques / Sciences de la nature
Troisième cycle	MSN 31 Pouvoir et résoudre des problèmes pour modéliser le plan et l'espace... Mathématiques	MSN 32 Pouvoir et résoudre des problèmes pour construire et structurer des représentations des nombres réels... Mathématiques	MSN 33 Résoudre des problèmes numériques et algébriques... Mathématiques	MSN 34 Mobiliser la mesure pour comparer des grandeurs... Mathématiques	MSN 35 Modéliser des phénomènes naturels, techniques, sociaux ou des situations mathématiques... Mathématiques / Sciences de la nature

Figure 3.1.2: Réseau des objectifs d'apprentissage pour les mathématiques selon le PER.

Dans cet exemple, la thématique *Modélisation* a une place particulière puisqu'elle traverse l'ensemble des autres thématiques (y compris en sciences de la nature qui possèdent par ailleurs trois autres thématiques).

À noter que chaque domaine dispose de sa propre logique quant à l'articulation entre les différentes thématiques. Ainsi, dans le cadre de ce projet, est-il nécessaire de penser la place et le rôle de chacune, qui plus est dans une logique évaluative.

- Une **progression des apprentissages** (voir figure 3.1.3) pour chaque *objectif d'apprentissage* (précisé de composantes); elle décrit, sur deux ans au degré primaire et par année au degré secondaire, *les connaissances et compétences qui doivent être abordées et travaillées en classe* (PER 2010). Exemple en mathématiques pour la thématique Espace au cycle 2 :

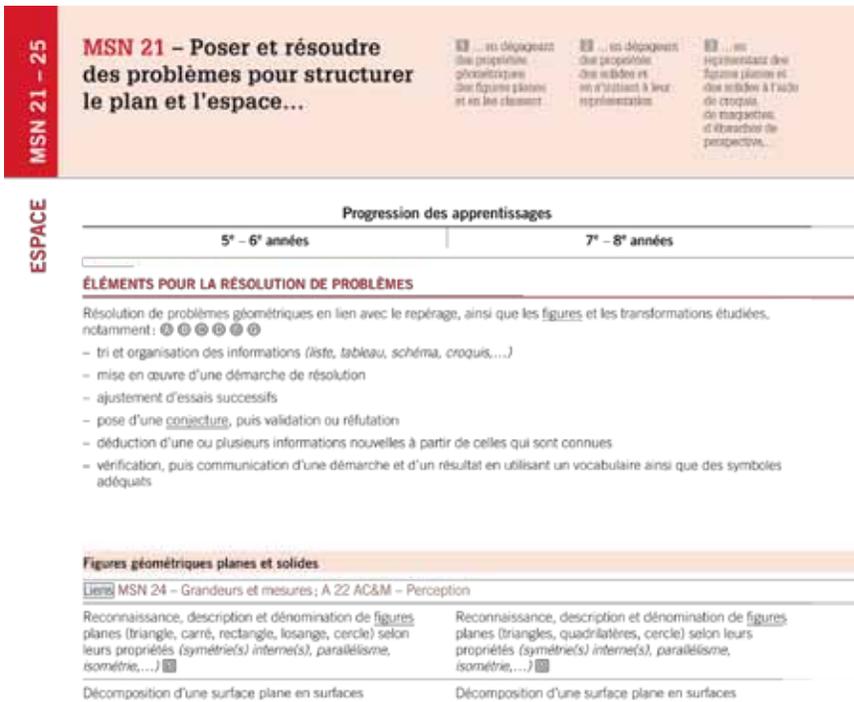


Figure 3.1.3 : Progression des apprentissages (extrait) associée à l'objectif d'apprentissage MSN 21 selon le PER

Relevons ici la trace de la thématique *Modélisation* qui apparaît au début de la *progression des apprentissages*; ceci est le cas pour chaque thématique et chaque cycle dans ce domaine.

- Des **attentes fondamentales** (voir figure 3.1.4), en lien avec la *progression des apprentissages*, décrivent les acquisitions qui sont nécessaires à la poursuite des apprentissages dans les disciplines concernées. Elles déclinent ce que tous les élèves doivent au moins maîtriser au cours, mais au plus tard à la fin du cycle. (PER

2010). Ainsi, les attentes fondamentales sont-elles présentées, pour l'enseignant, dans un but de régulation des apprentissages.

Pour faire suite à l'exemple ci-dessus, voici quelques attentes fondamentales pour la partie *Figures géométriques planes et solides* :



Figure 3.1.4: Attentes fondamentales (extrait) associées à l'objectif d'apprentissage MSN 21 selon le PER

Ajustement de la lecture du PER

Le danger d'une présentation type PER, est la tendance à une lecture aboutissant aux *attentes fondamentales*, laissant faussement croire qu'il s'agit là du but à atteindre avec les élèves. Or, les *attentes fondamentales* servent avant tout de repères pour l'enseignant dans le but de réguler voire de prioriser les apprentissages pour certains élèves.

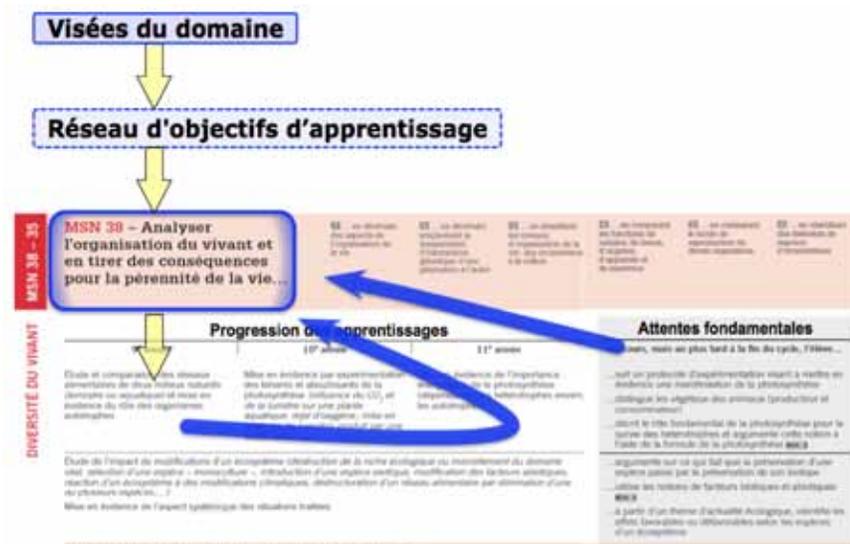


Figure 3.1.5: Imbrication des éléments structurels du PER et clé de lecture

Comme le schéma ci-dessus le présente, les *attentes fondamentales* ne sont pas la finalité de l'enseignement d'une discipline ; ce sont les *objectifs d'apprentissage* qui constituent cette finalité. Toutefois, les repères que fournissent les attentes fondamentales permettent de cibler les difficultés des élèves et, en ce sens, procurent des jalons non négligeables pour créer de l'évaluation.

Dans cette logique, toute évaluation – d'autant plus si elle est à visée sommative – prenant les *attentes fondamentales* comme objet d'évaluation ne devrait se faire qu'une fois le cycle terminé. C'est d'ailleurs ce que Gerard (2008) rappelle: *Une évaluation des acquis des élèves ne devrait JAMAIS avoir lieu avant que l'élève n'ait réalisé les apprentissages sur lesquels portera l'évaluation. (...) Une évaluation à fonction de régulation peut porter sur des acquis dont l'apprentissage est loin d'être terminé, mais il doit au moins être commencé!* (p. 48). Si ce discours semble plus qu'évident, nous avons pourtant pu constater toute la difficulté de coordonner le référentiel d'enseignement et le référentiel d'évaluation quand des contraintes temporelles et organisationnelles imposent des passations d'épreuves sommatives voire d'orientation en cours d'année scolaire, qui plus est avant la fin d'un cycle.

Nature des éléments du PER: quelques exemples

Dans chaque domaine du PER, les *visées du domaine* sont traduites, pour chaque cycle, en *objectifs d'apprentissages*. Ces derniers sont donc cohérents et complémentaires entre eux et relèvent de compétences visées pour chaque cycle. Pour chaque objectif, *la progression des apprentissages décrit (...) les connaissances et compétences qui doivent être abordées et travaillées en classe* (PER 2010) et précise donc dans quel sens l'objectif est travaillé. Enfin, les attentes fondamentales déclinent ce que tous les élèves doivent au moins maîtriser au cours, mais au plus tard à la fin du cycle et délimitent une « zone » qu'il s'agit pour tout élève d'atteindre durant un cycle donné.

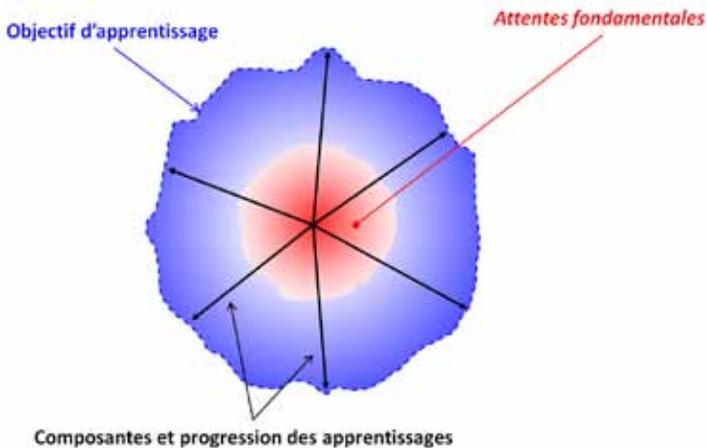


Figure 3.1.6: Représentation de l'articulation entre les éléments du PER décrivant un objectif d'apprentissage

Un exemple en mathématiques:

Au cycle 2, pour la thématique *Opérations*, l'objectif visé est *Résoudre des problèmes additifs et multiplicatifs* et, sous la *progression des apprentissages*, une partie est notamment réservée aux *Multiples, diviseurs, suites de nombres* où est précisé ce qui sera travaillé en classe avec les élèves. Les *attentes fondamentales* à ce sujet indiquent par exemple (...) *utilise les critères de divisibilité par 2, 5, 10, 100.*

En termes d'évaluation, plusieurs questions permettent de mesurer comment un élève utilise ces critères de divisibilité; cependant, les niveaux de difficulté et de complexité de ces questions varient. Il s'agit donc de se mettre d'accord sur le niveau d'exigence souhaité et un travail d'ajustement de ces questions est donc indispensable.

Par ailleurs, si chaque attente peut être mesurée à travers un ensemble de questions ajustées, cela demeure insuffisant pour indiquer comment un élève se situe par rapport à l'objectif *Résoudre des problèmes additifs et soustractifs*. Il est donc nécessaire de compléter cette première mesure d'une seconde portant sur l'objectif dans son ensemble.

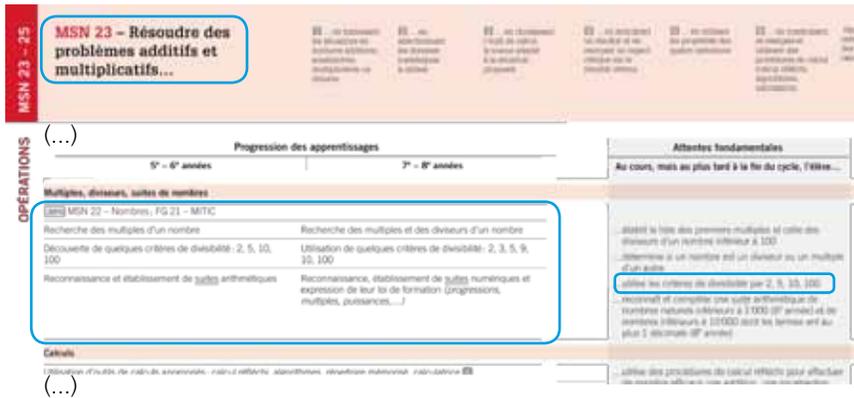


Figure 3.1.7: Exemple d'éléments de progression d'apprentissage et d'attentes fondamentales associés à un objectif d'apprentissage

Un exemple en français :

Toujours pour le cycle 2, un des objectifs visés en français est *construire une représentation de la langue pour comprendre et produire des textes (...)*. Parmi les progressions proposées, dans le sous-domaine orthographe grammaticale, l'une d'elles concerne l'accord du verbe avec le sujet dans les cas simples. L'attente fondamentale sur ce point est libellée comme suit : *(...) effectue, dans ses productions et dans les textes dictés, l'accord du verbe avec le sujet dans les cas simples, dans les cas où le sujet est inversé et lorsque le sujet est un pronom relatif.*

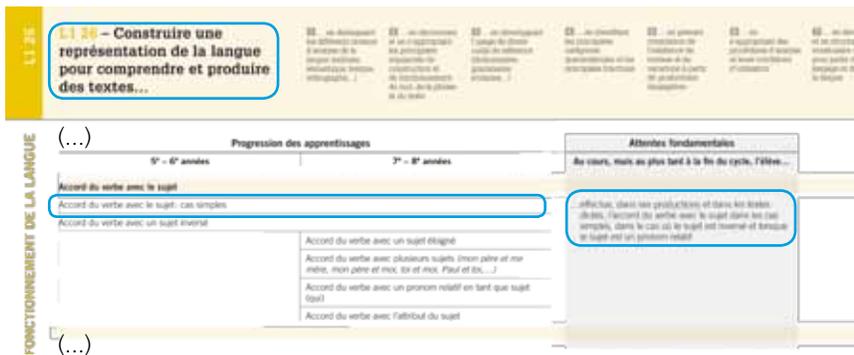


Figure 3.1.8: Exemple d'éléments de progression d'apprentissage et d'attentes fondamentales associés à un objectif d'apprentissage

Comme en mathématiques, il est possible d'imaginer un grand nombre de questions de difficultés et de complexités diverses pour mesurer une attente qui paraît pourtant relativement simple. De plus, il est bien différent d'envisager une mesure de cette attente si elle porte sur des phrases isolées ou sur la production d'un texte d'élève. Évidemment, la mesure d'une telle attente – même à l'aide d'un grand nombre de questions – est loin d'assurer la mesure de l'ensemble de l'*objectif d'apprentissage*, dont on voit bien la portée générale et le haut niveau de complexité.

Ces deux exemples soutiennent le constat que les *objectifs d'apprentissage* sont difficiles à évaluer. Les réflexions à ce sujet montrent que l'évaluation doit tenir compte des *progressions d'apprentissages* et des *attentes fondamentales* sans pour autant se limiter à ces dernières. Une mesure qui porterait uniquement sur les *attentes fondamentales* ne permettrait pas d'affirmer que les élèves ont bien développé la compétence visée inscrite comme *objectif d'apprentissage*.

3.2 Exemples d'évaluation en Communauté française de Belgique

Comme nous venons de le voir, prendre pour référentiel le PER implique de s'inquiéter de l'évaluation qui doit y être associée. Ces dernières années, de nombreux curriculums, programmes ou plans d'études formulés en terme de compétences ont vu le jour, du Canada à la Belgique. Les expériences des deux pays cités ci-dessus nous avertissent des difficultés de mise en œuvre d'un curriculum et offrent également des expériences intéressantes concernant l'évaluation. Nous nous sommes tout particulièrement intéressés aux évaluations belges qui posent la question d'une part d'une évaluation diagnostique externe et d'autre part d'une évaluation au service des enseignants basée sur un nouveau référentiel portant sur l'acquisition de compétences.

Des évaluations externes

Les évaluations externes sont des évaluations dont la conception et l'organisation sont confiées à des personnes extérieures à l'équipe éducative d'un établissement scolaire. Entre autres, les **évaluations externes non certificatives** en Communauté française de Belgique sont des évaluations à visée diagnostique et ont pour mission de fournir des outils pour les enseignants. Ces évaluations informent les équipes éducatives, mais également les responsables du système sur le niveau d'avancement des élèves. L'objectif premier est d'établir l'état des acquis des élèves (points faibles et forts). Les évaluations ont une visée diagnostique, les résultats qu'obtiennent les élèves n'affectent donc pas leur parcours scolaire.

La passation des épreuves est organisée sur une semaine (fin janvier) pour tous les élèves des classes de 4^e, 7^e et 10^e années HarmoS. Trois disciplines sont évaluées, à savoir la lecture-production d'écrits, les mathématiques et les sciences-éveil, avec pour référentiel les socles communs. La mise en place de ces évaluations nécessite l'implication de nombreux intervenants, inspecteurs, directeurs d'école, enseignants, élèves et animateurs pédagogiques.

Plusieurs conditions sont nécessaires pour réaliser des épreuves scientifiquement fiables et pour obtenir des informations valides :

- disposer d'un grand nombre de questions ;
- assurer que chaque compétence¹⁰ soit évaluée et cernée par plusieurs tâches ;
- concevoir les questions pour que les réponses puissent être codées de manière précise, univoque et standardisée ;
- maintenir la durée des corrections dans des limites acceptables avec pour conséquence que l'épreuve doit contenir un nombre important de questions fermées ;
- assurer des conditions de passation fortement standardisées (placer les élèves dans les mêmes conditions, réaliser les épreuves dans le même ordre, recevoir les mêmes consignes,...).

De plus, des essais préliminaires sont réalisés en début d'année scolaire auprès de 300 élèves pour assurer la fiabilité des épreuves.

Pour chaque question, le guide de codage spécifie les différents codes à attribuer selon les réponses (traitement particulier pour certaines questions). Des journées spéciales sont organisées par les inspecteurs pour corriger les épreuves. Le travail est collectif et réunit tous les enseignants. Les épreuves sont corrigées selon les consignes fournies dans le dossier de l'enseignant. Une fois les épreuves corrigées, tous les enseignants entrent les données dans une grille téléchargeable sur Internet. Les élèves y sont répertoriés selon leur numéro d'ordre attribué lors de la passation. Cette manière de procéder est facile à utiliser et permet de calculer immédiatement le pourcentage de réussite par élève et par classe.

Tous les élèves passent les évaluations. Puis, pour chaque niveau d'enseignement, les chercheurs constituent un échantillon représentatif de 2500 à 3000 élèves issus de 120 écoles. Seules les classes de l'échantillon doivent transmettre leur grille d'encodage qui sera traitée à l'université de Liège. Les résultats de l'échantillon sont analysés et font l'objet d'un rapport détaillé qui permet notamment à tous les enseignants de situer leur classe par rapport à l'échantillon. Ces résultats permettent également la conception d'un document comportant un certain nombre de pistes didactiques pour aider et soutenir les enseignants dans leur pratique.

Cette évaluation externe non certificative présente ainsi une expérience d'évaluation de compétences à visée diagnostique dont il serait possible de s'inspirer dans l'élaboration d'épreuves romandes communes.

Outils d'évaluation de compétences au service des enseignants

Le modèle d'évaluation développé par Rey et al. (2004) en Communauté française de Belgique est utilisé dans le cadre d'outils d'évaluation de compétences au service des enseignants du fondamental. Ce modèle est d'abord basé sur les trois degrés de compétence tels que les conçoivent les auteurs (voir point 2.2), pour mémoire :

- « *Compétence de premier degré* : savoir exécuter une action en réponse à un signal préétabli, après entraînement ; appellation : compétence élémentaire ou procédure ;

¹⁰ Compétence s'entend ici au sens large puisqu'elle fait référence à des compétences des trois degrés décrits par les auteurs (voir point 2.2)

- *Compétence de deuxième degré*: savoir choisir, parmi les procédures qu'on connaît, celle qui convient à une situation ou à une tâche non connue; appellation: compétence élémentaire avec interprétation (ou cadrage) de la situation;
- *Compétence de troisième degré*: savoir, parmi les procédures qu'on connaît, choisir et combiner celles qui conviennent à une situation ou à une tâche non connue et complexe; appellation: compétence complexe. » (p.33)

En correspondance inverse de ces degrés, sont proposés trois phases pour cette évaluation :

- La **phase 1** présente quelques tâches complexes demandant à l'élève de choisir et de combiner plusieurs connaissances (correspond aux compétences de 3^{ème} degré).
- La **phase 2** présente les mêmes tâches complexes avec toutefois une décomposition en tâches partielles; l'élève est ainsi guidé dans la réalisation de ces tâches complexes (correspond aux compétences de 2^{ème} degré).
- La **phase 3** présente des questions plus classiques portant sur les connaissances engagées dans les deux premières phases (correspond aux compétences de 1^{er} degré).

Le cadre de ces évaluations est bien plus souple que celui qu'impose une épreuve externe puisque *la passation peut être envisagée à n'importe quel moment du cursus scolaire à condition que l'apprentissage des savoirs et savoir-faire nécessaires à l'accomplissement des tâches proposées ait eu lieu. Il appartient à l'enseignant d'estimer quel est le meilleur moment pour réaliser ces évaluations.*¹¹ Ainsi perçoit-on tout l'intérêt d'évaluations diagnostiques externes pour l'enseignant qui dispose d'outils ajustés et calibrés lui permettant d'identifier le type de difficultés que ses élèves peuvent rencontrer.

Il est intéressant de rappeler l'historique du développement de ces épreuves en trois phases. Suite à l'adoption de *Socles de compétences* en 1999 par le gouvernement de la Communauté française de Belgique, les décideurs politiques ont pris le parti que soient proposées aux enseignants, à titre informatif, des *épreuves étalonnées construites en référence aux socles de compétences*. C'est dans ce cadre que les travaux de recherche menés par Rey et al. (2001) ont abouti à ce modèle évaluatif en trois phases, partant du constat que les compétences sur lesquelles devaient porter l'évaluation étaient peu maîtrisées par les élèves en général. La visée choisie pour ces épreuves était donc d'abord diagnostique et six épreuves ont été construites puis expérimentées auprès de 1200 élèves en tout, échantillon constitué auprès de 19 écoles et 45 classes.

Les principaux constats de cette vaste recherche sont les suivants (Rey et al. 2003, p. 125):

- La majorité des élèves réussit mal la phase 1, soit peine à résoudre des tâches complexes; les auteurs soulignent ainsi toute la pertinence de proposer une épreuve à visée diagnostique.

¹¹ *Guide d'accompagnement des outils d'évaluation de compétences au service des enseignants du fondamental*. Ministère de la Communauté française, administration générale de l'enseignement et de la recherche scientifique, service général du pilotage du système éducatif. 6 pages. www.enseignement.be

- La phase 2 obtient des résultats à mi-chemin entre la phase 1 et la phase 3, cette dernière étant, sans réel étonnement, la mieux réussie. À relever quelques doutes quand à la phase 2 qui impose une démarche de résolution à l'élève sans qu'elle soit celle qu'il aurait effectivement choisie pour résoudre la tâche complexe.
- L'ensemble du modèle est cohérent du fait des relations constatées entre les réussites à chacune des phases: un élève qui réussit la phase 1 réussit sans surprise la 2 et la 3.

En mai 2010, lors d'une journée d'étude organisée par l'IRD, V. Carette constate les difficultés d'élaboration de telles épreuves, déplore le temps important nécessaire à la passation et émet également quelques pistes d'amélioration de leur modèle, notamment l'exclusif support papier-crayon. Il rappelle la forte influence du vécu – dans et hors de la classe – des élèves dans l'appréhension d'une tâche complexe.

Ces modèles évaluatifs ont grandement inspiré notre travail et nous avons largement pris en compte les limites de ce modèle afin, comme nous allons le voir, de l'adapter à notre réalité.

3.3 Éléments pour une évaluation adaptée au PER

Bien que d'abord référentiel d'enseignement, le PER fournit, pour chaque cycle, des *objectifs d'apprentissage* et des *attentes fondamentales* (voir point 3.1); ces dernières sont écrites à des degrés de compétence divers, c'est-à-dire qu'elles traitent non seulement des connaissances à acquérir mais également de compétences plus larges. Parce que le PER est écrit comme un curriculum (voir point 2.1), la réflexion à mener dans notre projet nous conduit à nous demander, comme le firent Rey et al., quelles sont les caractéristiques d'une évaluation adaptée au PER. De fait, les *objectifs d'apprentissages* sont définis, contrairement à ce que l'intitulé pourrait laisser entendre, comme des compétences visées et nous semblent constituer ainsi l'objet principal d'évaluation. Afin de concevoir notre modèle d'évaluation, nous disposons également de *composantes* qui précisent dans quelle optique se développent ces compétences, ainsi que de compléments encore plus fins grâce aux *progressions d'apprentissage* et aux *attentes fondamentales*.

Nous présenterons donc comment, en s'inspirant du modèle d'évaluation en trois phases, nous avons caractérisé notre propre modèle d'évaluation, d'abord diagnostique, et quelles sont les possibilités qu'offre un support numérique dans ce cadre. Nous concluons ensuite sur quelques caractéristiques de notre modèle dans une perspective d'épreuve à visée sommative.

3.3.1 Un modèle d'évaluation en deux parties complémentaires

Créer une évaluation adaptée au PER implique de tenir compte de la nature des éléments qui structurent ce référentiel et mène donc au constat que la seule mesure de connaissances ne saurait suffire. La difficulté principale d'une évaluation qui ne porte pas seulement sur des savoirs se situe dans la prise en compte des compétences visées. En effet, évaluer un ensemble de connaissances et procédures isolées ne suffit pas pour assurer que l'élève soit capable de les mettre en œuvre dans une situation plus complexe, même si elle est ajustée au niveau attendu des élèves.

Les *Groupes de travail* et l'*Equipe de projet* ont tout d'abord exploré les diverses épreuves en leur possession, provenant de Suisse romande et de quelques pays francophones. Ils ont constaté que plusieurs d'entre elles tentent d'ores et déjà de dépasser les questions « classiques » portant uniquement sur les savoirs et tendent à mesurer, par des activités plus amples, des compétences de degré 2 voire 3 (voir point 2.2). Celles-ci nécessitent bien sûr de mobiliser des connaissances et des procédures mais exigent également de l'élève la mise en œuvre d'une démarche, d'une organisation dans son travail. C'est ce que nous désignerons sous le terme de *tâche complexe* par opposition à une *tâche spécifique* qui regroupe une série de questions ponctuelles portant sur des connaissances (savoirs, savoir-faire, ...).

Dans l'idée d'obtenir des informations d'une part sur les compétences annoncées dans le PER et d'autre part sur les attentes fondamentales exprimées à des niveaux taxonomiques divers, nous proposons une épreuve en deux phases :

- La première vise l'évaluation des ressources, c'est-à-dire de savoirs et savoir-faire (compétences de degré 1 et 2) qui peuvent être mobilisés dans la compétence plus vaste que l'on souhaite évaluer. Cette phase fusionne les phases deux et trois du modèle belge exposé précédemment (voir point 3.2) ; nous faisons le choix de proposer des questions courtes, fermées ou semi-fermées, voire des QCM.
- La seconde vise l'évaluation de compétences (compétences de degré 3) par des tâches complexes. Cette phase correspond donc à la phase une du modèle belge à ceci près que les situations proposées appartiennent à une discipline. La cohérence interne du PER nous permet de choisir quelques *objectifs d'apprentissage* et de considérer que cette seconde phase évalue également certains éléments inscrits dans les *attentes fondamentales*.

Les choix effectués de transformer les trois phases du modèle belge ainsi que l'ordre dans lequel elles sont proposées reposent sur les hypothèses suivantes :

- Le fait de disposer, dans le référentiel romand, de compétences disciplinaires nous impose de nous concentrer sur des **évaluations disciplinaires**, d'une part parce que l'attente institutionnelle nous semble forte d'obtenir des informations séparées à ce sujet, d'autre part parce que l'imbrication de connaissances et compétences disciplinaires à mobiliser nous semble une entrave pour détecter les sources d'erreur ; toutefois, nous travaillons notre modèle et développons en parallèle nos épreuves prototypiques en français et en mathématiques dans l'optique d'explorer les éventuels liens entre les deux.
- Les différents ouvrages cherchant à articuler connaissances et compétences tendent à prédire un échec de l'expression d'une compétence lorsque les connaissances à mobiliser sont déficientes. Cette idée est renforcée par le fort taux d'échec constaté à la réalisation des tâches complexes proposées par Rey et al. (2004) et le taux de réussite plus important dans leur phase 3. Sur la base de cette hypothèse, nous partons donc dans l'optique d'un éventuel redimensionnement, pour certains élèves, de la tâche complexe si constat est fait qu'ils ne maîtrisent pas suffisamment les connaissances nécessaires. En d'autres termes, nous envisageons la possibilité de ne pas proposer la seconde phase à des élèves ayant échoué à la première, voire de ne proposer qu'une fraction de la seconde épreuve ou encore de proposer une partie spécifiquement réalisée pour ces élèves. La possibilité de **varier l'ampleur de la seconde phase** nous semble d'autant plus

faisable que la première phase est présentée sur support numérique et permet d'automatiser un résultat individuel (voir point 3.3.2).

- Par ailleurs, envisager, dans une première phase, la mobilisation des ressources qui seront réinvesties dans la seconde, nous semble présenter un autre avantage : nous assurer que ces ressources soient stables. L'enjeu, dans la seconde phase, n'est pas de vérifier à nouveau que ces ressources soient effectivement au point, mais bien la façon dont l'élève va les articuler. En ce sens, la mesure obtenue lors de la seconde phase doit principalement fournir des informations sur la **combinaison** de ces ressources. S'il nous semble essentiel d'effectivement mesurer, en un premier temps, l'état de certaines connaissances, il ne faudrait pas que, dans un deuxième temps, celui-ci ait un effet bloquant sur la réalisation de la tâche complexe. Notons encore qu'afin de renforcer cet aspect mais également de gommer quelque peu l'impact de la culture préalable de l'élève en lien avec le contexte de la tâche complexe, nous avons proposé une activité préalable à la seconde phase. Ce choix a été largement discuté et l'épreuve prototype nous montrera si un écart important est constaté entre la première et la seconde phase dans l'utilisation de ces ressources.
- La phase 2 proposée dans le modèle belge nous semble peu satisfaisante du fait, notamment, que la procédure de résolution n'est pas forcément celle que l'élève aurait choisie (voir point 3.2). En outre, les questions fortement décontextualisées, donc très scolaires, proposées dans leur phase 3, nous semblent plus pertinentes dans le cas d'une remédiation et moins dans le cas d'une évaluation. C'est pourquoi le choix a été fait de constituer une unique phase regroupant des questions d'abord contextualisées mais qui ne sont pas encore de véritables tâches complexes, et comportant également quelques questions plus « classiques ».
- La volonté de réduire le temps nécessaire à l'évaluation (essentiellement dans le cadre d'une passation en classe complète) ainsi que le temps dévolu à ses corrections nous a menés à concevoir une partie informatisée (où chaque élève passe l'épreuve à son tour et à son rythme), réduisant ainsi la partie sur support papier. Dans la mesure où ce dernier peut être considéré comme un support privilégié des tâches complexes qui nécessitent, par nature, de l'espace pour explorer, organiser et composer les ressources mises en œuvre, la seconde phase est ainsi réalisée sur support papier. Toutefois, il est également envisageable, pour des élèves d'un certain âge, d'imaginer des tâches complexes intégrant d'autres supports ; nous y reviendrons.

Le temps d'apprendre...

Présentés par cycle, les *objectifs d'apprentissages* du PER, ainsi que nous l'avons mentionné précédemment, sont d'abord des compétences visées, comme l'expression *Résoudre des problèmes additifs* ou encore *Comprendre et produire des texte* le laissent largement penser. De plus, les *attentes fondamentales* font suite à la mention *Au cours, mais au plus tard à la fin du cycle, l'élève...*, ce qui signifie qu'il existe ici une véritable volonté de laisser le temps aux apprentissages de se développer et se stabiliser. Proposer des épreuves adaptées au PER nous impose donc de tenir compte du positionnement de ces éléments et nous interroge quant au moment à privilégier pour une passation.

Il est nécessaire de rappeler la forte disparité entre cantons, premièrement dans l'existence même d'épreuves externes cantonales, deuxièmement dans les années scolaires et le moment où peuvent apparaître ces épreuves et troisièmement dans l'ampleur et le taux de recouvrement de ces épreuves. Ainsi avons-nous envisagé le moment de passation de nos épreuves de façon à ne pas supplanter les épreuves en place – sachant qui plus est que certaines figurent dans les lois cantonales – et de ne pas alourdir la charge et le moment de passation.

Afin de laisser effectivement se faire les apprentissages et de tenir compte des contraintes temporelles de passation, nous avons privilégié une optique diagnostique en début de cycle: les compétences visées pour le cycle précédent devraient ainsi être suffisamment développées et les *attentes fondamentales*, du moins en ce qui concerne les ressources annoncées, devraient bénéficier d'une bonne stabilité. L'imbrication des *objectifs d'apprentissages* d'un cycle à l'autre dans le PER présente l'avantage de ne pas se retrouver en porte-à-faux en début de cycle, vis-à-vis des apprentissages qui y seront menés. Tenir compte, en partie, des apprentissages prévus dans le cycle suivant, dans une idée pronostique, permet à une telle évaluation d'avoir une fonction de régulation et les résultats, assortis de pistes didactiques, pourraient proposer des situations de remédiation tout en s'inscrivant dans le développement des compétences visées pour ce cycle.

Produire ainsi un message à destination direct de l'enseignant qui a charge de travailler avec des élèves dans l'année – voire les deux années – suivante(s), nous paraît faire sens et, espérons-le, permettre une bonne adhésion à ces mesures de la part des différents acteurs.

En résumé:

Le modèle d'évaluation que nous proposons se compose donc de deux parties, l'une constituée de *tâches spécifiques* et l'autre de *tâches complexes* (voir figure 3.3.1). La première partie permet de vérifier un ensemble de connaissances et de procédures, la seconde d'analyser comment l'élève les mobilise dans une situation donnée. Ce modèle suit ainsi de près la structure du PER, la première partie portant plus spécialement sur les attentes fondamentales, la seconde sur l'objectif visé. Une cohérence interne est également assurée pour l'ensemble de l'épreuve par la prise en compte des éléments constituant la première partie pour composer la seconde.

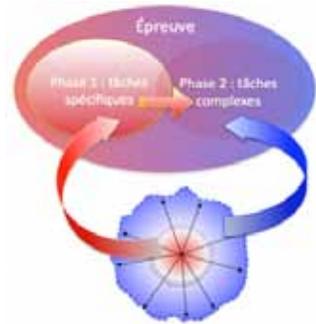


Figure 3.3.1: Représentation schématique de la prise en charge des éléments du PER dans l'épreuve en deux phases

3.3.2 Un développement possible: une épreuve adaptative

Comme nous l'avons évoqué, l'idée d'articuler deux parties dans une épreuve tentait de répondre à des contraintes curriculaires mais aussi organisationnelles. L'intérêt d'une épreuve informatisée, qui plus est adaptative, réside dans les nombreuses possibilités et informations **complémentaires** à une épreuve papier-crayon. Toutefois,

les réticences exprimées depuis plusieurs années concernant ce support sont importantes, en particulier quant au doute d'une maîtrise suffisante de l'outil par les élèves (de la manipulation de la souris à la familiarité même de l'outil).

Les principales raisons d'explorer un support informatisé tiennent, pour une épreuve à visée diagnostique, d'une part au fait que la partie papier-crayon doit se concentrer sur le strict nécessaire et d'autre part au fait qu'il s'agit d'examiner aussi finement que possible les difficultés de certains élèves, sans alourdir la passation pour les élèves qui n'en ont pas. Déjà en 1974, T. Nelson présentait sa première devise dans *Computer Lib* : *utiliser un ordinateur devrait toujours être plus simple que d'accomplir la même tâche sans ordinateur*. La question se pose en ces termes : que peuvent apporter les TIC que les épreuves papiers ne sauraient fournir – ou, à l'inverse, quelle seraient leurs limites – d'abord du point de vue de l'élève, ensuite du point de vue de l'institution ?

L'adaptativité en question

Préalablement, il est nécessaire de clarifier dans quelle perspective nous entendons le terme « épreuve adaptative ». Les expériences dont nous avons pris connaissance lors de ces travaux sont principalement *Stellwerk* et *Basis-check*. Ces deux modèles sont essentiellement conçus sur une distribution des items selon leur difficulté calculée par une TRI. Par conséquent, un premier constat, qui s'inscrit mal dans nos perspectives, est que la succession des questions est uniquement liée à des réussites ou échecs à une succession de questions dont l'ordre des sujets abordés – dans une discipline annoncée – est, *a priori*, aléatoire (Behrens, Vermot & Marc, 2011). Or, selon nous, viser un diagnostic implique d'obtenir des informations complémentaires sur certaines questions échouées, précisément au moment où l'élève vient de se confronter à une tâche. Le format des tâches que nous envisageons doit être souple, entendu par là que nous ne souhaitons pas cantonner l'élève dans des réponses strictement fermées : nous visons, au contraire, la possibilité pour lui d'essayer, de manipuler, de recommencer. Cette possibilité complémentaire présente le double avantage que l'élève est rarement en position de « sécher » sur une question – c'est-à-dire de ne pas répondre ou de répondre « au hasard » – et que la séquence des manipulations effectuées fournit un reflet assez clair du niveau de démarche de l'élève, entre démarche experte et démarche exploratoire.

Relevons que des développements importants ont lieu ces dernières années. Notamment la mise à disposition des enseignants, dans certains pays, d'outils de développement de tests en ligne. En particulier au Luxembourg qui propose un site et une application en ligne¹² permettant l'élaboration de tests par les enseignants qui le désirent. Les évolutions sont importantes et, en quelques années, les fonctionnalités proposées se sont élargies alors que nos premières explorations à ce sujet nous paraissaient peu satisfaisantes, le format des questions étant alors limité à des QCM, zones de texte ou menu déroulant.

Les possibilités de l'épreuve et des tâches

Sur support informatisé, la navigation dans l'épreuve entre diverses ressources (texte à lire, écoute de consigne, ...) peut non seulement être clairement préparée et conte-

¹² <http://www.tao.lu/html/>

nue, mais aussi prélevée pour chaque élève. De plus, il est possible de proposer aux élèves qui se trompent de revenir sur leur réponse, permettant ainsi l'observation des interventions des élèves sur leurs propres productions. Ceci reste envisageable sur papier mais la gestion et les interventions nécessaires pour ce faire sont vite considérées comme démesurées par les enseignants. Par ailleurs, la réponse à une erreur peut ne pas se limiter à sa seule correction mais être une tâche complémentaire, permettant une exploration affinée des causes de l'erreur. La succession des activités, avec des tâches réservées à certains élèves selon le niveau de réussite, est ainsi une réponse nouvelle au suivi des erreurs. Les possibilités pour les concepteurs s'amplifient donc de façon considérable dans la prise en compte de pistes d'explications relatives aux réponses erronées, selon des hypothèses didactiques sur les raisons d'un échec à une tâche.

Une autre option à explorer est l'apparition des informations constitutives d'une tâche. Les temps prélevés pour certaines étapes montrent la grande disparité dans l'appropriation des tâches par les élèves, nous rappelant la modestie que doivent conserver nos analyses *a priori* et nos interprétations dans les démarches supposées des élèves. La possibilité de mettre sous les yeux des élèves des objets dynamiques permet de questionner différemment leur compréhension et ouvre un champ d'exploration en termes d'évaluation. Par exemple, dans une évaluation en mathématiques – mais peut-être également dans d'autres domaines – la possibilité de manipuler des objets permet des observations nouvelles des démarches des élèves, avec une fiabilité que les interprétations effectuées jusqu'à présent sur les traces papier des élèves n'offraient pas. Cette possibilité, mais également des champs d'écriture, offrent secondairement un avantage incontestable par rapport au papier : l'impression pour tous les élèves de fournir un travail abouti, propre et sans trace de leurs essais ou erreurs, impression satisfaisante que seuls les très bons élèves connaissent habituellement sur papier.

Les informations prélevées et l'automatisation de leur traitement

Outre les réponses aux items, l'outil informatique permet de prélever un nombre important d'informations dont certaines sont de nature différente de celles prélevées dans une épreuve papier-crayon. Une partie introductive de l'épreuve peut permettre d'observer le comportement des élèves face à l'outil (voir point 4.2). Ensuite, il est possible de prélever le temps d'appropriation d'une tâche, son temps de réalisation, les manipulations effectuées, le temps de lecture d'un texte (en français p. ex.), l'écoute – ou non – de texte oraux. La difficulté de ce genre d'épreuve est alors non pas d'obtenir les informations mais de les traiter et d'en retenir les plus pertinentes.

Dans le cadre d'une épreuve adaptative, il est possible, comme évoqué précédemment, de rajouter à certains élèves des tâches complémentaires ; celles-ci sont proposées lorsqu'une tâche a été loupée et sont construites sur les hypothèses des causes de l'échec à l'item précédent. On peut également envisager de proposer des tâches complémentaires lorsque les élèves en réussissent une en particulier, avec l'idée de prospecter sur l'élargissement des connaissances mises en œuvre. La mise en évidence de difficultés récurrentes dans la réalisation des tâches permet de caractériser une sorte de profil de l'élève, résultat instantané des connaissances de l'élève pour ces éléments testés. Tout l'intérêt de programmer l'épreuve informatisée pour en extraire un « instantané » de chaque élève réside dans la possibilité de fournir rapidement une information à l'enseignant avec des pistes de remédiations tout à fait actuelles pour

l'élève – si tant est qu'il soit possible d'en définir. Une partie papier peut tout spécialement y faire écho; elle est alors à disposition des enseignants comme complément à l'épreuve papier-crayon que passent tous les élèves.

Les équipements TIC des ménages en Suisse

Il nous a semblé nécessaire, ici, de faire l'état des lieux des équipements informatiques des ménages en Suisse, eu égard à quelques réactions sceptiques que nous avons pu observer à la proposition d'une partie d'épreuve informatisée. S'il est vrai que, dans cette situation tout particulièrement, il ne saurait être pertinent de rajouter une difficulté liée à l'outil, il s'agit également de voir dans quelle mesure, vu l'évolution exponentielle de la place des nouvelles technologies dans notre société, ce support ne permettrait pas, au contraire, de rompre avec des approches évaluatives qui ont déjà montré leurs limites.

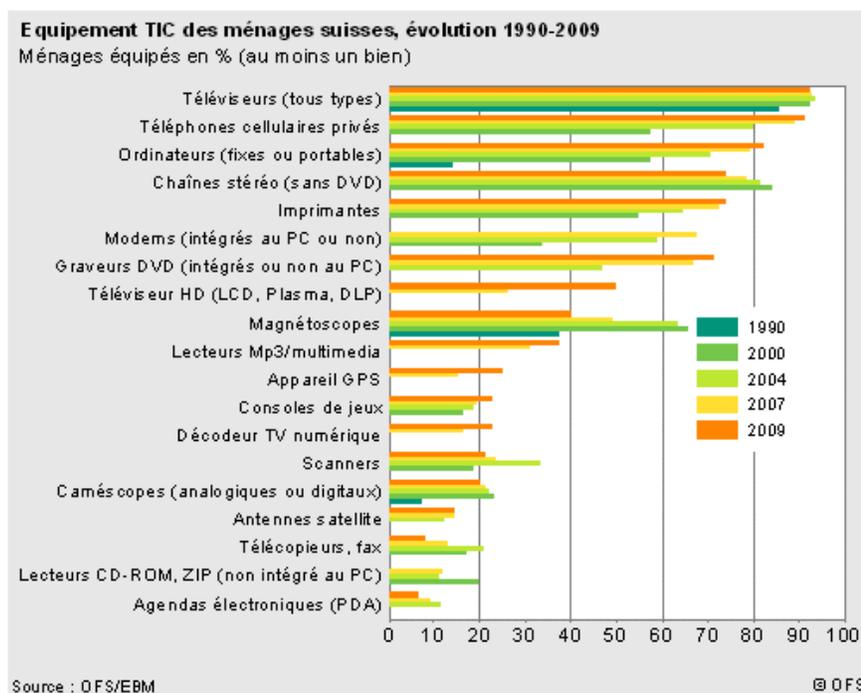


Figure 3.3.2a: Technologies de l'information et de la communication – équipement des ménages suisses entre 1990 et 2009

Les dernières données de l'Office Fédéral de la Statistique¹³ quant aux équipements TIC (technologies de l'information et de la communication) des ménages datent de 2008. Elles montrent une très forte augmentation de l'équipement en ordinateurs depuis 1990:

¹³ http://www.bfs.admin.ch/bfs/portal/fr/index/themen/16/04/key/approche_globale.indicator.30103.301.html/

Dans les années 90, l'équipement informatique des ménages en Suisse a augmenté très rapidement. La part des ménages munis d'un PC au moins a plus que quintuplé entre 1990 et 2009 (elle passe de 15% à plus de 82%). Alors qu'en 1990 un ménage sur 100 seulement était équipé de deux PC ou plus, cette proportion s'élève à 35% en 2009. La forte baisse des prix des PC ainsi que la popularité grandissante de l'internet dans la population suisse ont contribué de manière déterminante à l'augmentation marquée de l'équipement en ordinateur des ménages. Si le taux d'équipement est en lien avec le niveau de formation, on constate toutefois une moyenne de plus de 81% de l'ensemble des ménages qui sont équipés d'un ordinateur personnel (plus de 94% dans les ménages de milieu tertiaire).

Plus intéressant encore est l'équipement selon la taille des ménages. Dès qu'un ménage comporte trois personnes ou plus, le taux de ménages équipés d'ordinateurs dépasse 96%.

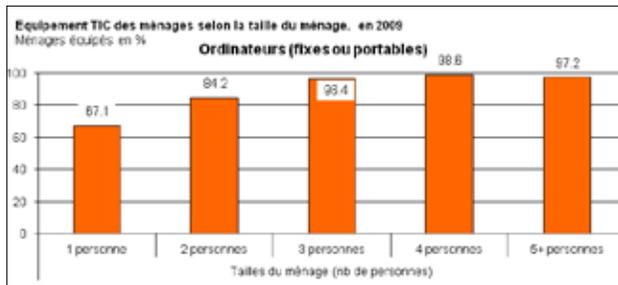


Figure 3.3.2b : Équipement des ménages suisses en ordinateurs fixes ou portables en 2008

« Le fait principal à retenir réside dans le haut niveau d'équipement des ménages de 4 personnes et plus, en particulier en ordinateurs personnels et en consoles de jeu. En général, la présence d'enfants dans le ménage apparaît ainsi comme un facteur clé dans la diffusion des nouvelles technologies au sein de la société. »

Mettant en perspective ces données qui datent déjà de quelques années, il nous a semblé intéressant d'explorer les possibilités d'une épreuve adaptative dans le cadre de nos travaux, examinant à cette occasion l'habitude des élèves à l'utilisation de l'outil. Comme nous le présenterons dans la partie 4.1, nous avons donc opté pour une première phase d'épreuve sur support informatique qui, pour une épreuve diagnostique, semble tout à fait adapté.

3.3.3 Un modèle adaptable à la visée évaluative

Le modèle en deux phases exposé sous 3.3.1 présente de nombreux avantages dont le premier est d'articuler les savoirs et les compétences que l'on souhaite mesurer, émanant d'un *objectif d'apprentissage* clairement identifiable dans le PER. Cette épreuve en deux parties permet également de pouvoir considérer plusieurs objectifs à la fois, correspondant à une évaluation à visée sommative, ou, au contraire, ne tenir compte que d'une partie restreinte de l'objectif, correspondant à une évaluation à visée diagnostique (voir les deux scénarios présentés sous 1.5).

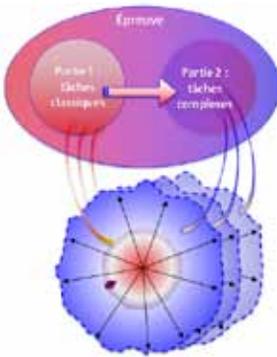


Figure 3.3.3a : Prise en compte de plusieurs objectifs à la fois : adapté à une évaluation sommative voire bilan

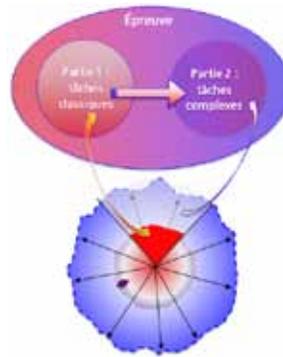


Figure 3.3.3b : Prise en compte d'une partie d'un objectif : adaptée à une évaluation à visée diagnostique

La figure 3.3.3a montre que cumuler plusieurs objectifs à évaluer dans une seule épreuve limite la finesse de la mesure : l'information obtenue sera plus globale et donc moins opérationnelle pour un enseignant qui souhaiterait remédier aux difficultés de certains de ses élèves. Inversement, ne prendre qu'une portion restreinte de l'objectif (figure 3.3.3b) fournira une information très pointue sur les objets d'apprentissage retenus, donnant des pistes précises pour remédier à des difficultés de l'élève, mais ne donnera pas une vision globale de l'atteinte des objectifs.

À l'heure actuelle d'une première mise en œuvre du PER, le développement d'épreuves à visée diagnostique permettrait d'explorer finement chacun des objectifs du PER et ainsi de constituer et d'ajuster de nombreuses tâches d'évaluation. Parmi celles-ci, et après passation de l'épreuve, quelques-unes s'avèreront particulièrement bien correspondre aux *attentes fondamentales*. Une épreuve à visée diagnostique permet de créer un grand nombre de questions fines sur un objectif particulier, constituant les premiers éléments d'une banque d'items. Cette démarche donne l'occasion d'échanges sur l'interprétation des éléments du PER, accompagnant ainsi la mise en œuvre de ce référentiel. Privilégier d'abord le développement d'épreuves à visée diagnostique présente l'avantage de laisser ouverte la possibilité de développer des épreuves à visée sommative.

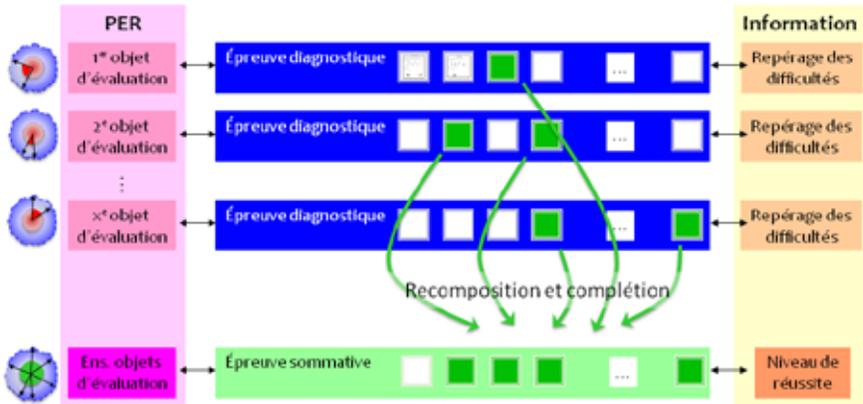


Figure 3.3.3c: Utilisation des questions dans une autre épreuve

Après quelques années et selon la volonté politique, il sera alors possible de créer des épreuves ayant d'autres visées (sommative ou encore bilan) (figure 3.3.3c). En effet, plusieurs des questions évaluatives développées pour des épreuves diagnostiques pourront être réinvesties dans une nouvelle épreuve qu'il s'agira de caractériser par un travail complémentaire (choix des objectifs, calibrage de l'épreuve, ...). À cette occasion, les éventuels ajustements du PER, notamment aux standards, pourront également être pris en compte. La réflexion devrait s'étendre au statut à donner aux *objectifs d'apprentissage* du PER, formulés en termes de compétence, que l'élève développe au cours du cycle. Dans l'idéal, une synthèse des informations obtenues devrait s'ancre autour de ces *objectifs d'apprentissages*, notamment pour un public moins spécialisé comme les parents voire pour un public plus large.

4. Élaboration d'un prototype d'évaluation en deux phases

Le modèle d'évaluation que nous proposons se compose donc de deux parties, l'une constituée de *tâches spécifiques* et l'autre de *tâches complexes*. La première partie permet de vérifier un ensemble de connaissances et de procédures, la seconde d'analyser comment l'élève les mobilise dans une situation donnée. Ce modèle suit ainsi de près la structure du PER, la première partie portant plus spécialement sur les *attentes fondamentales*, la seconde sur l'*objectif d'apprentissage*. Une cohérence interne est également assurée pour l'ensemble de l'épreuve par la prise en compte, à l'intérieur de la seconde partie, des éléments constituant la première partie.

Ce modèle présente l'avantage d'articuler les savoirs et les compétences que l'on souhaite mesurer, sur la base d'un *objectif d'apprentissage* clairement identifiable dans le PER. La question de savoir dans quel ordre proposer les phases aux élèves ne peut être éludée dans un premier temps. *A priori*, par cohérence avec les pratiques d'apprentissage et comme le font d'ailleurs Rey et al., il semblerait préférable de partir de la tâche complexe, qui est l'élément principal visé par l'évaluation puisque relatif à la compétence, puis de proposer des questions plus ponctuelles sur les ressources; les informations au sujet de ces dernières devraient ainsi permettre de comprendre les éventuelles lacunes manifestées par les élèves et à même d'expliquer les difficultés rencontrées dans la tâche complexe (voir point 3.3).

Nous n'avons cependant pas procédé de cette manière. En effet, en commençant par les questions sur les connaissances et démarches, nous nous sommes donné pour objectif de parvenir à établir des « profils » d'élèves, l'épreuve adaptative offrant la possibilité de construire des cheminements différenciés en fonction des réponses justes ou fausses des élèves. L'idée était, initialement, d'explorer la possibilité de créer ces « profils » et, sur cette base, de voir comment les exploiter directement dans l'épreuve papier, par exemple en présentant des parties complémentaires ciblées en fonction des réponses obtenues à la première phase. Nous n'avons toutefois pas été aussi loin dans notre expérience et nous contenterons donc, dans ce premier essai, de lier ces « profils » aux résultats obtenus pour la tâche complexe et de proposer quelques pistes de réflexion reprenant ces premières idées.

En l'état de nos connaissances sur les liens entre tâche complexe et ressources mobilisées pour la résoudre, il apparaît que l'ordre de passation des phases est finalement moins important que de tenter de mieux appréhender quels sont ces liens et ce qu'ils disent des connaissances et compétences des élèves. Engagés dans une démarche exploratoire, nous ne pouvons donc qu'émettre des hypothèses à ce propos et, sur la base des traitements statistiques de nos données, constater dans quelle mesure elles se vérifient ou non; une fois ces questions débroussaillées, il nous sera alors peut-être possible de clarifier l'influence de l'ordre de passation de chaque phase en fonction de l'information visée.

Le modèle d'évaluation en deux phases que nous avons retenu semblant pertinent, l'essai d'une mise en œuvre devait permettre d'expérimenter le dispositif en gardant à l'esprit les questions suivantes :

- Le dispositif est-il pertinent pour les deux disciplines ?
- Comment cerner le champ évaluatif en se référant au PER ?
- Comment caractériser les items de la première phase, les tâches complexes de la seconde, et comment les articuler ?
- Quelle plus-value peut apporter un test informatisé pour cette première phase et comment créer des étapes supplémentaires pour un test adaptatif destiné à identifier les sources d'erreurs des élèves ?
- Dans quelle mesure le dispositif proposé convient-il à l'enseignement-apprentissage en classe ? Les questions sont-elles ajustées aux élèves ?
- Existe-t-il des liens d'une part entre les deux phases de test et d'autre part entre le test de français et celui de mathématiques ? Quel traitement statistique est le mieux à même de révéler ces liens ?
- Peut-on extraire des données permettant de cibler des difficultés des élèves et de proposer des pistes de remédiation ?

Il a été choisi d'expérimenter le dispositif dans une perspective diagnostique pour la fin du cycle 1, en prévoyant un test auprès d'élèves de début de 5^e année, entre septembre et octobre. Une dizaine de classes se sont prêtées à cette expérience.

4.1 Les options retenues dans l'élaboration des épreuves

Dans les deux disciplines, le premier travail a tout d'abord consisté à cerner l'objet d'évaluation en référence au PER. Un *objectif d'apprentissage* du cycle 1 dans chacune des disciplines a été retenu comme objet premier de l'évaluation. La *progression des apprentissages* et les *attentes fondamentales* associées nous ont servi à repérer les savoirs concernés et à établir le niveau de difficulté et de complexité des items et des tâches. Une des grandes difficultés réside dans l'ajustement des tâches complexes. *A priori*, nous évitons de présenter des situations par trop inédites qui tendraient, comme évoqué au point 2.1 et constaté dans les résultats de l'évaluation en trois phases de Rey et al. (voir point 3.2), à produire de l'échec dont on retire finalement peu d'information (Crahay, 2006).

De plus, il nous a semblé que l'*objectif d'apprentissage* du cycle 2, dans la même thématique (voir structure du PER point 3.1), permettait d'apporter des informations utiles en lien avec d'éventuelles questions plus pronostiques que diagnostiques, afin de renseigner la suite de certains apprentissages. De telles questions peuvent soit servir de tâches complémentaires, soit être insérées dans des tâches complexes. Grâce à la cohérence du référentiel d'un cycle à l'autre, ces questions n'apparaissent pas en décalage par rapport à celles proposées en lien avec le cycle 1.

La constitution de l'épreuve informatisée se fait sur la base des éléments considérés comme mobilisables dans la tâche complexe. Dans une épreuve en deux phases, une question essentielle apparaît à laquelle nous tenterons de répondre grâce aux analyses

effectuées : **peut-on présupposer la transférabilité des ressources d'une phase à une autre ? et peut-on extraire une information** sur les possibilités de transfert des connaissances des élèves ?

Le temps de passation de l'épreuve a été défini avant tout en fonction de l'ampleur de chaque phase du test. Chacune forme un tout cohérent ; la première explore les éléments de savoirs qui apporteront des informations pertinentes à la seconde et, la seconde explore d'avantage la compétence visée à travers des tâches complexes. Globalement, après élaboration des contenus de ces deux parties du test, la passation informatisée dure en moyenne trente minutes et la passation papier-crayon varie fortement selon la discipline, celle de l'épreuve de français prenant davantage de temps à cause, essentiellement, d'une longue mise en situation mais aussi de la présence de deux volets, un de compréhension, l'autre de production au sein de la même situation. Notons encore que pour chaque épreuve informatisée, une page de familiarisation à l'outil est présentée aux élèves ; elle comprend des manipulations telles que sélectionnement par clic, glissement d'objet, remplissage d'un champ textuel, ajustement du son.

4.2 Évaluation en deux phases en français à la fin de cycle 1

Le travail accompli pour le français a été enrichi à la fois des discussions qui ont eu lieu au sein des deux groupes de travail et de celles propres au groupe de français qui a, tout comme le groupe de mathématiques, régulièrement œuvré seul. Des décisions relatives au modèle lui-même, que nous voulions commun aux deux disciplines, ont été prises dans les séances communes aux deux groupes, alors que celles touchant aux choix spécifiques de la discipline français ont été prises dans le groupe disciplinaire.

4.2.1 *Caractéristiques de chaque phase pour le français*

Rapidement, les membres du sous-groupe français ont retenu deux points qui leur semblaient importants pour la phase 1 : tout d'abord, ils ont voulu éviter un schéma d'épreuve conventionnel qui fait que ce qui touche à la structuration de la langue (grammaire, orthographe, vocabulaire et conjugaison) soit exclusivement dans une phase, ici la phase 1, et ce qui touche à la compréhension et à la production se trouve entièrement dans l'autre phase, soit la phase 2. Cela pourrait laisser penser en effet que dans un premier temps il est indispensable de surmonter les tâches techniques pour, dans un second temps, accéder à la compréhension. En réalité, l'enseignement-apprentissage des techniques de lecture et celui de la compréhension de texte se déroulent conjointement. De plus, il existe des techniques, des stratégies, susceptibles de construire la compréhension de texte, et qui nous paraissaient avoir leur place dans la phase 1 (elles correspondraient, par exemple, aux compétences de degré deux chez Rey et al.). En d'autres termes encore, il s'agit de se demander ce qui doit figurer dans une des phases qui expliquerait ou pourrait être relié à l'autre phase. Dans cette perspective, et parce que le cycle 1 est essentiellement consacré à l'entrée dans l'écrit, il a paru important de mettre l'accent sur la lecture et de trouver des traces de la compréhension écrite dans la phase 1 déjà. Quelles sont les connaissances, les procédures ou stratégies de compréhension que les élèves doivent maîtriser en fin de ce cycle pour les aider à lire dans des tâches complexes ?

Entrée dans l'écrit signifie que lecture et écriture sont travaillées conjointement. Le sous-groupe a donc pensé la phase 2 dans cette optique, envisageant une tâche complexe qui lierait lecture et écriture. Il fallait encore que la phase 1 apporte des informations sur les bases de l'écriture, ce que les élèves doivent maîtriser pour aborder la production dans une tâche complexe.

4.2.2 Quelques choix dans l'élaboration de l'épreuve adaptative

Une fois ces points tranchés, il a fallu d'une part choisir un texte et des questions relatives à des stratégies de lecture ou des connaissances à acquérir pour le savoir-lire, d'autre part mettre en relation ces questions et les attentes fondamentales du cycle 1 touchant à la lecture et à l'écriture mais aussi au fonctionnement de la langue.

À propos des stratégies de lecture, nous nous sommes référés à l'ouvrage de Cèbe et Goigoux, *Lector & Lectrix* (2009), qui propose précisément de travailler un certain nombre de stratégies pour aider les élèves du primaire à lire des textes narratifs. L'ouvrage n'étant pas prévu pour de jeunes élèves du premier cycle, il n'a été pris en considération que partiellement. Ayant en outre pour but d'entraîner chez les élèves des habiletés de compréhension de l'écrit, il présente une certaine logique dans les activités didactiques proposées. Il n'aurait donc pas été pertinent de prendre des tâches de cet ouvrage pour en faire des questions d'évaluation. Néanmoins, cet ouvrage a le mérite de faire le tour des principales stratégies de lecture que les élèves doivent s'approprier pour aborder des textes narratifs et il nous paraissait intéressant d'en tenir compte dans une certaine mesure pour bâtir notre épreuve. Chacune des stratégies est travaillée dans une séquence d'enseignement-apprentissage : construire une représentation mentale du texte ; reformuler ; intégrer de nouvelles informations au fur et à mesure de la lecture du texte ; répondre à des questions ; faire des inférences, chercher à lire entre les lignes ; s'interroger sur le narrateur, les personnages, ce qu'ils disent, à qui, dans quel but, etc. Cette liste offrait, à nos yeux, des possibilités de questionnement permettant de vérifier la présence de l'une ou l'autre de ces stratégies pour la compréhension du texte proposé.

Nous avons examiné les épreuves dont nous disposions, issues des cantons, d'autres pays et aussi des travaux d'HarmoS. C'est finalement une épreuve HarmoS qui a été choisie pour constituer la base de notre phase 1. En effet, plusieurs membres du groupe français avaient participé à l'élaboration des épreuves – en particulier en lecture – dans le cadre du projet sur les standards de formation pour la L1. Le texte choisi est adapté à des élèves de fin de 4^e année. Chacune des questions renvoie à la fois à un aspect de compétence, selon les définitions du modèle de compétence élaboré par le consortium L1 d'HarmoS, et à un niveau de difficulté, selon les résultats obtenus sur un échantillon d'élèves suisses. Il était en outre possible de lier certaines d'entre elles aux stratégies travaillées dans *Lector & Lectrix*.

Il s'agissait encore – et surtout – de faire correspondre ces questions aux éléments du PER : un objectif d'apprentissage et ses composantes, des progressions d'apprentissage et les attentes fondamentales. Nous avons finalement retenu les éléments suivants :

COMPRENDRE ET PRODUIRE L'ÉCRIT

Nous sommes partis de l'objectif d'apprentissage suivant: *Lire et écrire des textes d'usage familial et scolaire et s'approprier le système de la langue écrite (objectif d'apprentissage de français L1 11-12 du PER)*, et des attentes fondamentales qui s'y rattachent.

<i>Attentes fondamentales: Au cours, mais au plus tard à la fin du cycle, l'élève...</i>	
	choisit le support écrit adapté à son projet
	identifie une syllabe, un phonème dans un mot
	repère dans un texte une phrase, un mot, une lettre
	lit et identifie des mots en prenant en compte ses connaissances du contexte, du code et de la correspondance phonème-graphème
	comprend le sens global de textes courts, courants ou littéraires, pourvus d'un contenu, d'une structure, d'une syntaxe et d'un vocabulaire d'usage familial
	trouve un mot dans un outil de référence créé dans le cadre de la classe* (lexique, affiche, liste de mots, etc.)
	lit un poème ou une comptine et l'interprète
	comprend le sens global (thème et idée principale) d'un petit texte narratif et nomme les personnages principaux
	écrit, à l'aide d'outils de référence, une ou plusieurs parties d'un texte qui raconte (environ 50 mots) sur des thèmes abordés en classe, en produisant des phrases syntaxiquement correctes et en respectant la segmentation lexicale
	comprend une consigne contenant une action décrite en termes courants

* Les passages du texte en gris concernent des éléments de l'attente qui ne peuvent concerner une épreuve externe ou qui ne peuvent être pris en compte que partiellement.

FONCTIONNEMENT DE LA LANGUE

Nous avons également pris en considération l'objectif d'apprentissage relatif au fonctionnement de la langue, à même de fournir des indications sur les ressources nécessaires à la compréhension et à la production de texte: *Observer le fonctionnement de la langue et s'approprier des outils de base pour comprendre et pour produire des textes (objectif d'apprentissage de français L1 16 du PER)*; à cet objectif d'apprentissage se rattachent les attentes fondamentales suivantes:

<i>Attentes fondamentales: Au cours, mais au plus tard à la fin du cycle, l'élève...</i>	
	GRAMMAIRE
	délimite dans des phrases simples (sans complément de phrase) des groupes de sens selon les fonctions de base (sujet et groupe verbal)
	ORTHOGRAPHE
	recherche des mots dans les documents de référence de la classe pour les orthographier correctement
	orthographie quelques mots de grande fréquence (thèmes travaillés en classe)

	énonce les 26 lettres de l'alphabet dans l'ordre
	classe des mots selon l'ordre alphabétique (lettre initiale différente)
	accorde le nom avec le déterminant dans un groupe nominal simple (pluriel en -s)
	VOCABULAIRE
	utilise un vocabulaire adéquat dans ses propres productions
	utilise ses connaissances du système lexical pour donner du sens à ce qu'il lit
	CONJUGAISON
	conjugue oralement au présent les verbes du type « chanter » + être, avoir et aller en utilisant les pronoms (je, tu, il, elle, on, nous, vous, ils, elles)
	DU TEXTE À LA PHRASE ET DE LA PHRASE AU MOT
	identifie les concepts de texte, de phrase, de mot, de lettre
	segmente une phrase écrite en mots
	reconnait et nomme toutes les lettres

Sur cette base, mais aussi en accord avec les choix faits (tenir compte des stratégies de lecture, faire une épreuve adaptative diagnostique), des questions complémentaires ont été créées pour parfaire l'épreuve envisagée, en particulier pour mesurer les attentes fondamentales liées au fonctionnement de la langue. Plusieurs problèmes se sont présentés au groupe au moment de réaliser l'épreuve, parmi lesquels :

- l'organisation de l'épreuve et l'ordre des questions : en compréhension de texte, nous avons décidé de commencer par les questions les plus faciles (au vu des résultats d'HarmoS) pour aller vers les plus difficiles ;
- le choix de questions : il devait, à nos yeux, permettre de tracer un chemin adaptatif à la fois pour les élèves éprouvant des difficultés particulières et pour ceux réussissant facilement les questions les plus difficiles ;
- le choix des distracteurs à l'intérieur de certaines questions : il nous a paru important que les distracteurs nous fournissent des informations sur la nature des réponses obtenues (types de difficulté par exemple) : nous avons donc fait des hypothèses interprétatives pour chacun d'eux ;
- les liens à définir avec l'épreuve de la phase 2 : il nous fallait choisir des questions sur des objets susceptibles d'une reprise à un niveau de complexité plus grand dans la tâche de compréhension et de production de texte de la phase 2 ;
- l'accès au texte pour des lecteurs peu habiles : nous avons prévu un retour au texte pour les questions qui auraient pu le nécessiter. Le retour au texte nous paraît dire quelque chose des stratégies de lecture des élèves. Nous avons également prévu l'écoute des consignes, afin d'éviter qu'elles interfèrent trop sur la capacité de répondre aux questions de compréhension du texte ou qu'elles fassent perdre beaucoup de temps aux élèves ayant quelque difficulté à lire. Dans un premier temps, nous avons envisagé l'accès au texte en trois étapes graduées : la lecture silencieuse par l'élève lui-même, la lecture du texte par l'enseignant, l'histoire du texte racontée avec des mots familiers. Le but de cette façon de procéder aurait été de voir si c'était le contenu du texte sous sa forme écrite qui posait problème ou plutôt sa lecture. Trop compliqué à mettre en place pour une épreuve externe,

nous avons finalement renoncé à ce scénario. De même, le groupe de travail s'est demandé dans quelle mesure il aurait été utile pour les élèves les plus en difficulté de disposer, dans certains cas au moins, de la définition ou de l'explication de certains mots (en fonction du contexte), sachant combien ces élèves buttent souvent sur le sens de mots difficiles à comprendre pour eux. Là encore, vu la complexité pour construire un tel dispositif (notamment avec les moyens à disposition dans cette étude exploratoire et en particulier à l'intention de si jeunes élèves), il a renoncé à le mettre en place; il n'en demeure pas moins que l'idée mériterait d'être poursuivie dans des évaluations ultérieures...;

- le choix des attentes fondamentales à mesurer, toutes ne pouvant être prises en considération : une analyse de chacune d'elles a dû être menée pour procéder à ce choix;
- l'association des questions ou de groupes de questions à d'éventuelles pistes de remédiation à proposer aux enseignants : idéalement, nous aurions voulu, déjà au moment de l'élaboration des questions, pouvoir prévoir les remédiations possibles en cas d'échec de la part des élèves.

La question du genre de texte à retenir pour la partie compréhension de l'épreuve s'est posée aux membres du sous-groupe français. Nous avons opté pour un conte, entrant dans le regroupement de genres *narrer*; si l'on examine les attentes fondamentales touchant à la compréhension de genres de textes précis pour le cycle 1, on voit que le conte et sa compréhension tiennent une place de choix tout au long du cycle. Il s'agit pour les élèves de s'en faire une idée globale, de comprendre qui sont les personnages et ce qu'ils font, ainsi que les événements qui se déroulent dans l'histoire. Les questions qui ont accompagné le texte (*Tibili le petit garçon qui ne voulait pas aller à l'école*¹⁴), ont donc porté sur ces éléments: une première question sur le choix du résumé correspondant le mieux au texte (parmi 4 propositions); une question d'interprétation d'un passage du texte; une question sur l'identification du genre; une question sur la compréhension d'éléments d'un dialogue; une question sur les renvois anaphoriques dans trois phrases du texte; une question de compréhension d'informations explicites du texte.

Les questions plus spécifiquement rattachées au fonctionnement de la langue ont été créées (ou reprises en partie du test d'HarmoS) par les membres du groupe pour permettre la vérification des attentes fondamentales liées à l'objectif d'apprentissage cité ci-dessus. Il s'est agi d'une question sur la segmentation de phrases en mots; une question sur le décodage de mots et la distinction de phonèmes à l'intérieur du mot; une question sur le repérage de phrases dans un court texte continu; une question sur l'identification et le classement de mots d'une même famille; une question sur la reconnaissance de pronoms de conjugaison et leur mise en lien avec le verbe conjugué correspondant.

Le plan de cheminement des questions, ci-dessous, illustre l'enchaînement des questions et montre aussi la place accordée aux questions complémentaires.

¹⁴ D'après *Tibili le petit garçon qui ne voulait pas aller à l'école*. Marie Léonard & Andrée Prigent. (2001). Paris : Magnard Jeunesse.

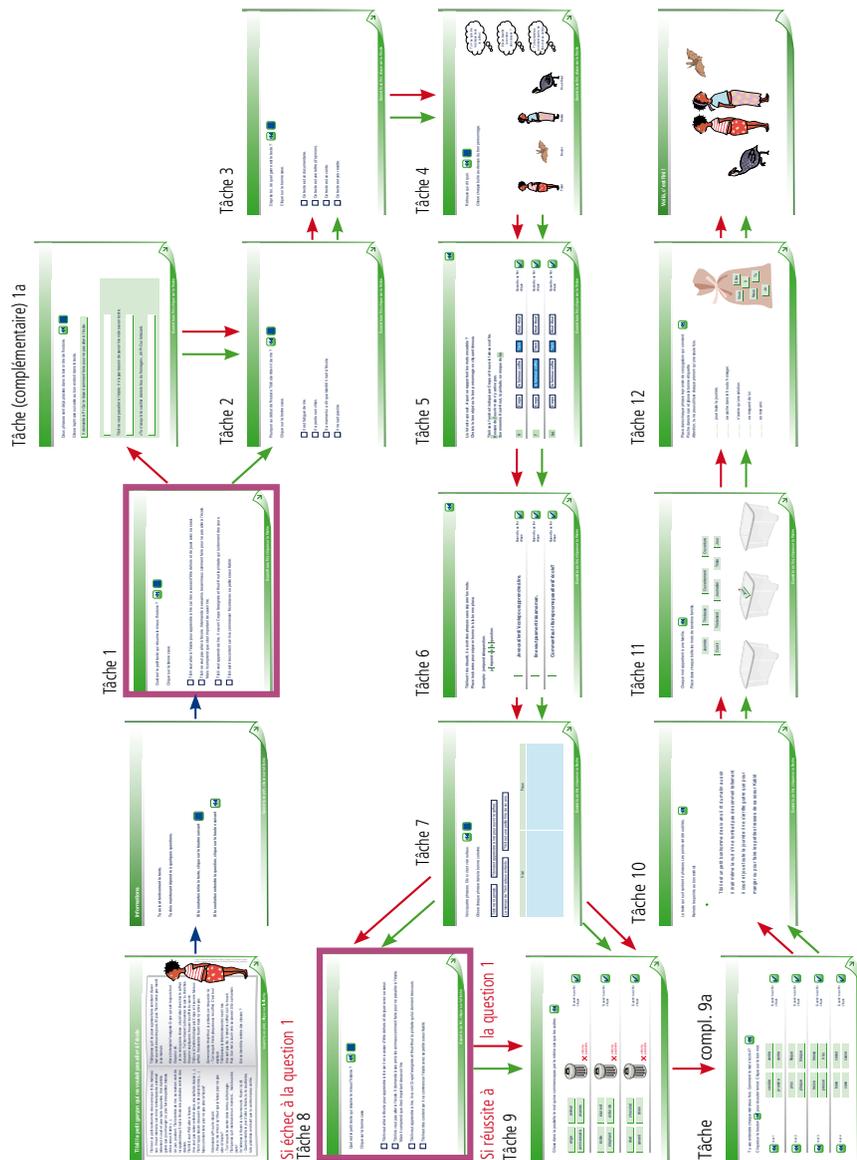


Figure 4.2.2: Succession des tâches spécifiques de l'épreuve adaptative français (se référer à l'annexe 8.3.)

→ l'élève réussit la tâche → l'élève ne réussit pas la tâche

Nous avons prévu deux questions complémentaires, mais aussi de placer une même question – la première, considérée comme facile dans le cadre d'HarmoS – à deux endroits du test: au tout début et au milieu. L'hypothèse a été la suivante: bien que considérée comme facile, certains élèves pourraient néanmoins échouer la tâche demandée; peut-être qu'en y revenant plus loin, après une série de questions sur le

texte et éventuellement plusieurs retours au texte, seraient-ils à même de finalement la réussir. Ce même questionnement à deux reprises nous a donc semblé fournir des informations intéressantes d'un point de vue diagnostique.

À cette première question, nous avons joint une question complémentaire pour les élèves en échec. Il s'agissait de placer deux propositions du texte au bon endroit par rapport à deux phrases données; cet exercice devait montrer si les élèves comprenaient l'ordre de certains événements du conte. Il nous semblait que cette façon d'envisager l'ensemble du texte devait être plus facile que de trouver le résumé le plus pertinent pour rendre compte de l'ensemble du texte. L'autre question complémentaire a concerné une question sur le fonctionnement de la langue relatif au lien phonème-graphème. Si les élèves échouaient cette question, ils en avaient une, que nous considérons comme plus facile, leur demandant de reconnaître un mot entendu parmi quatre mots proposés par écrit.

Dans la partie consacrée aux résultats des épreuves, nous verrons dans quelle mesure les choix et les hypothèses accompagnant cette épreuve informatisée à la fois s'avèreront pertinents et procureront des résultats intéressants pour le diagnostic.

4.2.3 *Quelques choix dans l'élaboration de l'épreuve papier*

Lors des discussions en groupe de travail, la question de la signification de la tâche complexe d'une part et celle de son contenu d'autre part se sont posées. Que veut dire tâche complexe pour des élèves de 3P? Quelle part donner à l'aspect inédit de cette tâche? Celle-ci ne doit-elle pas avoir des liens avec le programme? Le souci de fournir des informations utiles à l'enseignant pour poser un diagnostic est également apparu dans les discussions du groupe. Les choix effectués ont amené à caractériser la tâche complexe de manière différente de celle de l'équipe de Rey; inscrite dans la discipline, elle ne peut pas être considérée comme interdisciplinaire; de plus, elle n'est pas inédite, dans la mesure où elle permet des liens à la fois avec l'objectif d'apprentissage choisi et avec les attentes fondamentales retenues. Elle se rattache également aux propositions du moyen d'enseignement *S'exprimer en français*, à la disposition des enseignants de toute la scolarité obligatoire pour travailler, sous forme de séquences didactiques, les productions orale et écrite.

En tenant compte de l'*objectif d'apprentissage* (L1 11-12 *Lire et écrire des textes d'usage familial et scolaire et s'approprier le système de la langue écrite*), le groupe de travail a tenu à proposer une tâche complexe relative à la fois à la compréhension et à la production de texte. Le PER a aussi incité à faire le choix d'un genre de texte sur lequel faire porter la tâche. Nous avons choisi la lettre d'opinion (regroupement de genres *argumenter*) qui ne donne lieu à une attente fondamentale qu'à partir du cycle 2, bien qu'abordable déjà au cycle 1. Et ce dans une perspective diagnostique, pouvant inclure une part pronostique qui nous paraît utile à l'enseignant et son enseignement pour la suite de l'année en cours. Actuellement encore, nous nous demandons s'il n'aurait pas été tout de même préférable de rester dans le genre retenu dans l'EA, le conte, qui aurait permis davantage de liens entre les deux épreuves (ce que nous avons envisagé dans un premier temps). Nous y reviendrons lors de l'analyse de nos résultats.

Plusieurs problèmes restaient encore à résoudre. Pour parvenir à un diagnostic fructueux, la correction envisagée devait fournir aux enseignants des informations sur les

difficultés des élèves à entrer dans la tâche, à utiliser des procédures adéquates et à mobiliser les ressources nécessaires pour réussir la tâche. Comment récolter de telles informations ? Si nous avons choisi de faire passer l'EA en premier, c'était afin d'obtenir des informations éclairantes pour l'épreuve papier. Dans cette optique, faisant l'hypothèse que des profils d'élèves allaient pouvoir être définis à partir des résultats obtenus, nous avons, dans un premier temps, imaginé créer des documents d'aide à la résolution de la tâche pour les élèves en difficulté, selon leur profil. L'idée était ambitieuse et nous paraît encore pertinente aujourd'hui, même si nous n'avons pas pu la réaliser, faute de disposer des profils à temps.

Un autre problème auquel nous avons été confrontés a été d'éviter de placer de si jeunes élèves devant l'impossibilité de résoudre la tâche demandée et de fournir une page blanche, inutile pour un diagnostic ! C'est pourquoi, en nous inspirant des procédures mises en place dans les évaluations genevoises du primaire¹⁵ que nous trouvions intéressantes, nous avons prévu une activité préalable permettant aux élèves une entrée dans la situation de communication et leur fournissant un certain nombre d'informations facilitant l'accès à la tâche complexe. Si cette activité préalable apporte des avantages évidents, elle comporte également des risques ; en effet, elle peut être considérée comme une phase d'apprentissage qui va avoir des influences sur l'évaluation réalisée ensuite et, par exemple, courtcircuiter ce que nous aurions pu apprendre sur les ressources mobilisées par les élèves dans une tâche complexe.

Une dernière remarque concerne la prudence à conserver avec les résultats obtenus dans une tâche complexe. Il serait dangereux de considérer ces résultats en termes de réussite, échec, juste ou faux, sachant d'une part que l'objectif d'apprentissage est d'un haut niveau de compétence et ne pourra qu'être approché et non atteint (on apprend à écrire des textes de genres divers toute sa vie), et d'autre part la tâche complexe ne porte que sur un texte, un genre, et donc conduit à relativiser la réussite ou l'échec lié à une situation donnée. Dans une perspective diagnostique, il nous semble plus intéressant de tenter de mettre en lien les différentes informations obtenues pour aider à faire progresser les élèves, en particulier ceux qui présentent des difficultés dont les spécificités devraient être identifiées précisément grâce aux épreuves passées.

Comment se présente la tâche complexe construite pour notre épreuve papier ?

Structure de l'épreuve

L'épreuve papier de français comporte un grand nombre de documents, ceux destinés à l'enseignant et ceux destinés à l'élève (cf. 8.3.4).

Matériel de l'enseignant : dans un cahier, le déroulement pas à pas de l'épreuve est présenté, précisant comment procéder pour la compréhension et la production de textes, ainsi que pour la relecture prévue le lendemain.

Activité préalable à la compréhension : elle comprend plusieurs étapes. L'enseignant annonce le projet d'écriture : les élèves vont devoir répondre à la lettre de leur camarade Paul. Ensuite il ouvre une discussion en classe sur la consommation de bonbons, afin d'entrer dans le thème de la lettre. Des arguments pour et contre sont

¹⁵ Épreuve cantonale 2004. Français 1. Département de l'instruction publique. Enseignement primaire genevois.

donnés oralement. Enfin, à l'aide d'étiquettes fournies en annexe et que l'enseignant a préparées, il entreprend un travail sur les arguments. Disposant des étiquettes (chacun ayant lu la sienne), les élèves sont appelés alors à classer les arguments en pour et contre. Le classement se fait au tableau noir sur lequel les étiquettes sont posées. À la fin de l'activité, les étiquettes sont retirées du tableau noir.

Tâche de compréhension écrite : l'enseignant distribue à chacun la lettre de Paul et un questionnaire correspondant, placés sur une double page. Les élèves commencent par une lecture silencieuse de la lettre. Ensuite, après avoir donné des informations générales les concernant, ils sont invités à répondre individuellement aux questions du questionnaire. Une grille de correction et un tableau de spécification (présentant les composantes de la tâche, les objectifs poursuivis et les questions correspondantes) sont fournis en annexe à l'enseignant, à sa disposition pour procéder à la correction des questionnaires s'il le désire. Ces éléments doivent lui permettre de comprendre ce qui est visé par l'épreuve et la nature des difficultés rencontrées éventuellement par certains élèves.

Activité préalable à la production : afin de s'assurer que tous les élèves ont compris le contenu de la lettre de Paul avant de réaliser la production de texte qui leur est demandée, une vérification collective est réalisée. En un premier temps, les élèves sont amenés à comprendre quels sont les indices qui permettent de reconnaître une lettre. Ensuite, les questions du questionnaire sont reprises une à une.

Tâche de production écrite : les élèves reçoivent un dossier qui va leur permettre de réaliser la tâche. À l'intérieur se trouve la lettre de Paul. L'enseignant lit la lettre puis explique ce qu'il s'agira de faire. Pour que les élèves n'oublient pas la tâche, il écrit au tableau les consignes suivantes :

Avant d'écrire mon texte :

1. *Je me rappelle que je dois écrire une partie de lettre.*
2. *Je choisis mon opinion.*
3. *Je prépare dans ma tête deux arguments qui défendent mon opinion.*
4. *Je cherche une expression qui montre bien que c'est ma propre opinion que je donne.*

Les élèves découvrent également un début de lettre à compléter. Les éléments présents sont répertoriés et identifiés. De plus, ils disposent d'une feuille volante comprenant une liste de mots, dont beaucoup ont été utilisés lorsque, collectivement, ils ont cherché des arguments pour et contre le fait de manger des bonbons. Durant l'écriture de leur texte, les élèves peuvent demander à l'enseignant de compléter la liste d'un ou plusieurs mots qu'ils ne savent pas orthographier.

L'enseignant, quant à lui, reçoit une feuille d'observation et une explication des rubriques où consigner les observations à faire; cette feuille lui sert à noter les demandes des élèves à propos de l'épreuve. Le but de cette observation est de pouvoir offrir à l'enseignant la possibilité de garder des traces concernant des procédures des élèves. Cette observation a lieu au moment de la production de texte et au moment de sa relecture avec corrections éventuelles de la part des élèves.

Tâche de relecture du texte et corrections : le jour suivant, les élèves reprennent leur texte et reçoivent un guide de relecture de leur texte. Ce guide est lu collective-

ment et l'enseignant s'assure ainsi que chacun a compris ce qui lui est demandé. Les corrections éventuelles que les élèves voudraient apporter à leur texte se font à l'aide d'un crayon de couleur.

Dans les annexes du document-maitre se trouve encore la grille de critères de correction de la production écrite ainsi que la grille de spécification. S'il le désire, l'enseignant peut ainsi corriger les productions de ses élèves.

Que cherche-t-on à mesurer ?

Nous nous référons à Dolz & Schneuwly (1996), pour lesquels la notion de capacités langagières évoque *les aptitudes de l'apprenant requises pour la production d'un genre dans une situation d'interaction déterminée: s'adapter aux caractéristiques du contexte et du référent (capacités d'action), mobiliser des modèles discursifs (capacités discursives) et maîtriser les opérations psycholinguistiques et les unités linguistiques (capacités linguistico-discursives* (p. 58). Nous tentons d'étendre ces aptitudes à la compréhension également.

Dans ce sens, et sur cette base, nous avons construit des tables de spécification précisant les éléments mesurés dans la tâche complexe. L'influence des auteurs genevois cités plus haut et celle du plan d'études genevois est marquée, en particulier dans la terminologie utilisée. Ainsi, il faut entendre par composantes, les éléments qui viennent préciser, en le rendant plus opérationnel, le contenu de l'objectif-noyau, au centre de l'enseignement-apprentissage défini dans le plan d'études genevois de 2000: cet objectif-noyau est une *compétence essentielle de haut niveau privilégiée dans le cadre d'un cycle d'apprentissage et d'une discipline donnée. Il organise un réseau d'objectifs plus spécifiques en leur donnant structure et cohérence* (Les objectifs d'apprentissage de l'école primaire genevoise 2000). Les composantes s'articulent autour des caractéristiques langagières suivantes: le contexte, le contenu, la planification et la textualisation. S'y ajoute la révision (que nous avons prise en considération dans notre évaluation) comme travail sur le texte en étapes successives dans le but d'améliorer la production ou la compréhension du texte. Les objectifs spécifiques sont des *objectifs d'apprentissage compris dans les composantes de l'objectif-noyau qui précisent l'objectif-noyau encore plus finement que les composantes. Dans les attentes de fin de cycles, les objectifs spécifiques servent de critères pour l'évaluation de l'objectif-noyau ou de ses composantes* (Les objectifs d'apprentissage de l'école primaire genevoise, 2000).

Dans notre cas, précisons que nous n'avons pas introduit la planification dans la colonne des composantes. Il est clair que cela est propre à cette évaluation-ci et ne saurait être généralisé à d'autres évaluations ultérieures. Nous proposons même que la planification soit systématiquement présente dans les évaluations de tâches complexes en français. Comme précédemment, nous présentons les définitions suivantes tirées de *Les objectifs d'apprentissage de l'école primaire genevoise* de 2000:

« **Le contexte** renvoie à une caractéristique essentielle de la langue, de toute langue, à savoir qu'elle s'adapte à la situation dans laquelle elle est utilisée et dont les paramètres principaux sont l'énonciateur, le destinataire et le but poursuivi par l'action langagière.

Le contenu désigne les idées que toute action langagière véhicule. Produire un texte suppose que l'on élabore un contenu en fonction du thème et du contexte de production et que l'on sélectionne parmi l'ensemble des informations celles qui sont en adéquation avec la situation de communication.

La planification souligne que le contenu d'un texte doit être organisé en fonction du genre textuel produit, en fonction également du contexte dans lequel le producteur de texte se trouve (...)

La textualisation recouvre l'ensemble des procédés de mise en mots. Différents problèmes sont ici à traiter: syntaxiques, sémantiques, lexicaux, orthographiques... ainsi que l'emploi des unités linguistiques telles qu'organisateur (connexion), reprises et temps des verbes (cohésion) et modalités (modalisation). »

Si nous décrivons en détail ces éléments et faisons référence au plan d'études genevois, c'est que nous voyons une parenté forte entre ce document et le PER. La parenté concerne aussi bien la configuration des compétences et leur déploiement disciplinaire que la conception de la langue en termes d'activités langagières autour des genres textuels.

Les tables de spécification précisent ce que l'on cherche à mesurer. Voici celle qui concerne la compréhension écrite.

Composantes	Objectifs spécifiques	Questions
Contexte	Identifier la situation d'énonciation – l'énonciateur – le destinataire – but du texte	1 2 7a et 7b
Contenu	Identifier les différents points de vue Distinguer un argument d'une opinion	3 et 6 4
Textualisation	Repérer une expression introductive d'un argument (« parce que »)	5

La tâche complexe introduit une situation de communication, un contexte de production dont les éléments doivent être compris des élèves. La composante *contexte* est donc spécifique et donne des informations complémentaires à celles de l'épreuve adaptative.

L'option que nous avons prise de proposer dans la tâche complexe un texte d'un genre différent de celui de l'épreuve adaptative ne permet pas de retrouver des stratégies de celle-ci présentes dans la première.

Il reste que nous pouvons faire des hypothèses par rapport à des aspects du fonctionnement de la langue qui nous paraissent essentiels pour parvenir à lire et à comprendre un texte: les élèves doivent être capables de:

- Identifier une syllabe, un phonème dans un mot.
- Repérer dans un texte une phrase, un mot, une lettre.

- Lire et identifier des mots en prenant en compte ses connaissances du contexte, du code et de la correspondance phonème-graphème.

Il y aura donc un lien à établir entre les résultats aux questions de l'EA portant spécifiquement sur ces objectifs et le résultat en compréhension à la fois de l'EA et de l'épreuve papier.

Pour la tâche de production écrite, la table de spécification est la suivante :

Composantes	Objectifs spécifiques	Questions
Contexte	– saisir les caractéristiques d'une lettre	FP1 et FP5
Contenu	– émettre une opinion	FP2
	– produire des arguments cohérents	FP4
Textualisation	– utiliser des expressions modales de prise en charge	FP3
	– produire des énoncés syntaxiquement corrects	FP6
	– respecter les contraintes orthographiques	FP7 et FP8

Les liens qu'il est possible de faire entre épreuve papier et épreuve adaptative concernent la segmentation des phrases en mots et la reconnaissance des phrases d'un texte, correspondant à la production d'énoncés syntaxiquement corrects. L'utilisation des expressions modales de prise en charge (faire précéder son opinion de « je pense que », « à mon avis », etc) pourrait, dans une certaine mesure seulement, renvoyer à la bonne correspondance du pronom de conjugaison avec le verbe. De plus, le respect des contraintes orthographiques répond aux attentes fondamentales du PER suivantes :

- rechercher des mots dans les documents de référence de la classe pour les orthographier correctement ;
- orthographier quelques mots de grande fréquence (thèmes travaillés en classe).

Ainsi que :

- trouver un mot dans un outil de référence créé dans le cadre de la classe (lexique, affiche, liste de mots, etc.).

En français, l'objectif d'apprentissage touchant à la compréhension et à la production écrite est très vaste ; il en est de même pour celui relatif au fonctionnement de la langue. Comment les aborder sans créer des épreuves à large spectre ? De plus, dans le cadre d'une étude exploratoire, nous avons été tentés de considérer une large palette d'objets à mesurer, précisément parce qu'il est d'emblée difficile d'établir des liens entre eux, en particulier entre les ressources (savoirs et procédures) et la tâche complexe. Les données recueillies ont été nombreuses, voire pléthoriques, comme nous le verrons plus loin ! Bien des incertitudes et des hypothèses demeurent.

4.3 Évaluation en deux phases en mathématiques fin de cycle 1

Comme pour le travail en français, l'élaboration de l'évaluation en deux phases s'est d'abord faite conceptuellement avec les deux groupes disciplinaires puis les choix touchant spécifiquement aux contenus mathématiques se sont faits au sein du groupe mathématiques. Des mises en commun régulières avec le groupe français ont permis d'interroger plusieurs choix et d'observer l'évolution du modèle liée aux spécificités de chaque discipline.

4.3.1 Répertoire des éléments du référentiel à évaluer

Le groupe mathématiques s'est penché sur les thématiques qui structurent les mathématiques dans le PER, à savoir *Espace, Nombres, Opérations, Grandeurs et mesures* et *Modélisation*. Dans une visée diagnostique, il a paru primordial, en fin de cycle 1, d'explorer tout particulièrement où en étaient les élèves relativement à la construction du système décimal. En effet, de nombreux constats effectués chez des élèves en difficulté mathématique, et ce jusqu'au cycle 3, mettent en évidence une mauvaise appréhension et compréhension du nombre. C'est ainsi que nous avons choisi de créer une évaluation sur ce thème :

NOMBRES

Poser et résoudre des problèmes pour construire et structurer des représentations des nombres naturels (objectif d'apprentissage de mathématiques MSN 12 du PER)

<i>Attentes fondamentales: Au cours, mais au plus tard à la fin du cycle, l'élève...</i>
Résout des problèmes numériques en faisant appel à une ou plusieurs des composantes suivantes:
– choix et mise en relation des données nécessaires à la résolution
– utilisation des propriétés du système de numération
– vérification de la pertinence du résultat
– communication de la démarche (oralement)* et du résultat

Dénombrer une collection d'objets inférieurs à 50 par comptage organisé
Constitue une collection ayant un nombre donné d'objets inférieur à 50
Écrit en chiffres et récite (de manière fluide pour les nombres à deux chiffres) des séquences de la suite numérique jusqu'à 200
Compte de 10 en 10 jusqu'à 100, à partir d'un nombre donné
Passe du mot-nombre (oral) à son écriture chiffrée et inversement
Compare, ordonne, encadre, intercale des nombres inférieurs à 100

* Les passages du texte en gris concernent des éléments de l'attente qui ne peuvent concerner une épreuve externe ou qui ne peuvent être pris en compte que partiellement

MODÉLISATION

Représenter des phénomènes naturels, techniques ou des situations mathématiques (objectif d'apprentissage de mathématiques MSN 15 du PER)

Ce dernier *objectif d'apprentissage* est considéré comme transversal aux autres thématiques mathématiques et se concrétise directement à travers leur description en *progression d'apprentissage* et en *attentes fondamentales*. Ceci est d'ailleurs visible dans le premier bloc d'attentes du tableau ci-dessus (*Résout des problèmes numériques...*). Le groupe a également exploré l'idée de proposer une épreuve en *Logique* traitant principalement de l'objectif *Modélisation*. Toutefois, les difficultés à caractériser des pistes de remédiation semblaient, pour un premier essai, plus importantes et devoir bénéficier d'une expérience préalable dans un domaine mieux connu des enseignants.

La réflexion concernant les tâches spécifiques et complexes s'est donc orientée sur les *Nombres*, avec comme fil conducteur la construction du système décimal. Cette épreuve diagnostique en mathématiques vise donc à détecter les difficultés de certains élèves à comprendre la signification des chiffres composant un nombre pris dans le champ numérique des nombres naturels jusqu'à 200.

Très rapidement, un consensus a émergé au sein du groupe concernant les tâches complexes en question dans la mesure de compétences mathématiques. Dans le respect de la définition adoptée dans nos travaux concernant une compétence (voir chapitre 2.1), le groupe a retenu l'idée qu'une tâche complexe en mathématiques doit être un problème, associé à un contexte suffisamment proche de l'élève pour que ce dernier puisse l'appréhender, et menant à la mobilisation et à la combinaison des connaissances mathématiques ainsi que des composantes de résolution telles que celles décrites dans le référentiel PER.

4.3.2 Quelques choix dans l'élaboration de l'épreuve adaptative

Les principaux choix concernant l'épreuve adaptatives ont été les suivants :

- inscrire les questions dans un contexte donnant du sens aux tâches proposées et tentant de réduire l'impact d'un registre verbal et représentatif par trop variant ;
- exploiter le support informatique autant que possible en recherchant une plus-value aux fonctionnalités associées aux tâches et en évitant de créer des tâches qui sembleraient par trop destinées à un support papier ;
- couvrir globalement les *attentes fondamentales* relatives à l'*objectif d'apprentissage* sans s'y réduire ;
- élaborer des tâches spécifiques qui ne soient pas uniquement sur des contenus mais qui proposent également de « petits » problèmes, sans être déjà de véritables tâches complexes.

Ce dernier choix en particulier mérite développement. Les principales hypothèses à ce sujet sont d'abord que ces « petits » problèmes fournissent des informations sur les premières *attentes fondamentales* et constituent, *a priori*, des tâches d'application¹⁶

¹⁶ Selon le niveau d'assimilation de certaines connaissances, des tâches d'application peuvent constituer de véritables situations pour certains élèves.

(proches de celles proposées en phase 2 du modèle de Rey et al.); ensuite que les questions plus ciblées, décontextualisées, mettant en œuvre une connaissance spécifique, seraient à réserver à des diagnostics plus fins encore, ou être utilisées, par exemple, comme tâches complémentaires, voire encore être proposées comme remédiation pour certains élèves.

Globalement, il a paru essentiel au groupe de travail de présenter des tâches susceptibles de fournir une information sur la représentation que se font les élèves de l'organisation décimale des nombres. La majorité des tâches spécifiques proposées dans l'épreuve adaptative sont donc centrées sur la composition et décomposition de nombres en dizaines et unités; elles exigent parfois un dénombrement, des comparaisons et un classement de nombres.

Pour certaines tâches qui ne seraient pas réussies par l'élève, des questions complémentaires sont proposées; il peut s'agir soit d'une question qui reprend et allège la première question d'une partie des concepts, soit d'une question qui réinterroge l'élève sur le même objet en lui annonçant qu'il a fait une erreur, parfois avec une aide supplémentaire. Un essai a été fait de rajouter une question complémentaire pour les élèves ayant réussi une tâche en particulier; ceci c'est essentiellement fait dans un but exploratoire et quelque peu pronostique.

La figure 4.3.2 ci-après présente la succession de tâches spécifiques qui constituent l'épreuve adaptative. Les tâches complémentaires en cas d'échec à la tâche précédente sont encadrées en rouge et la tâche complémentaire en cas de réussite est encadrée en vert:

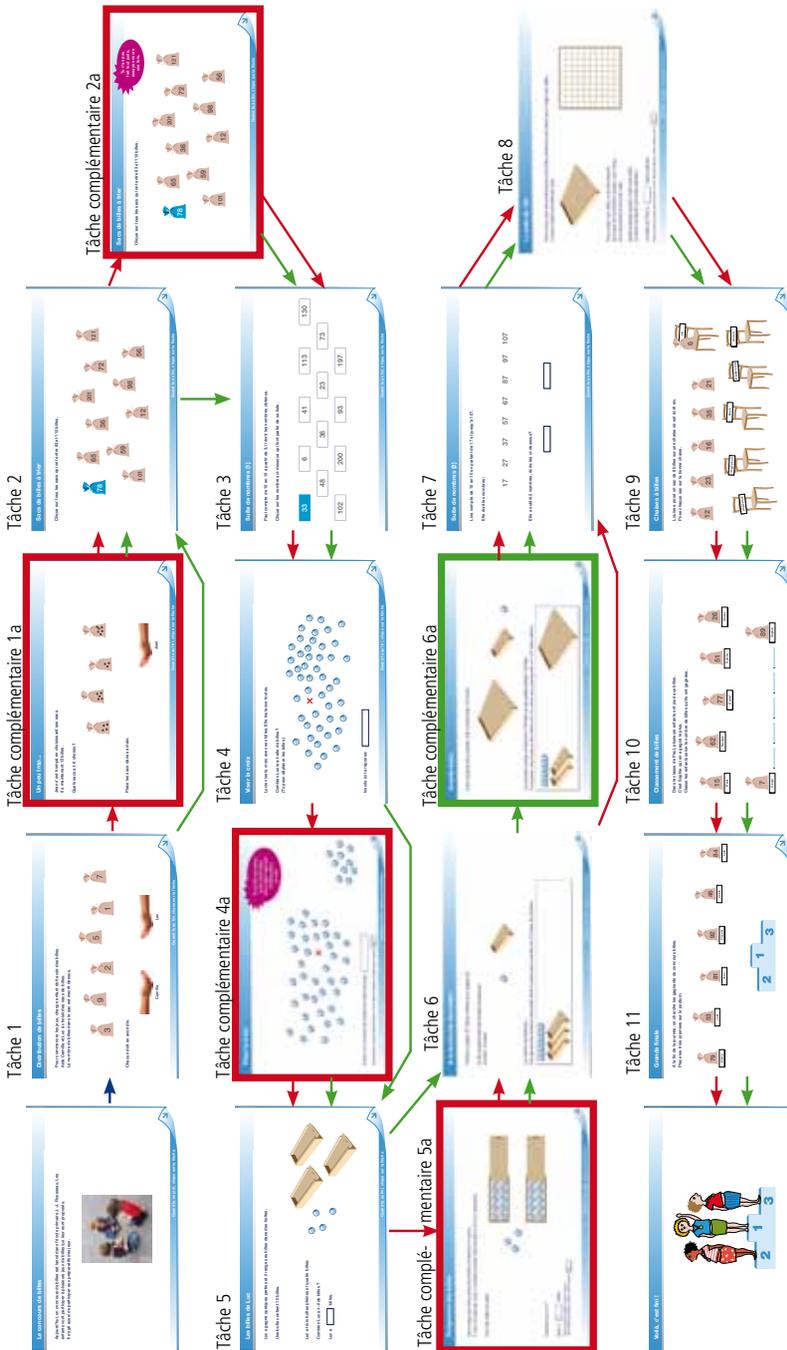


Figure 4.3.2: Succession des tâches spécifiques de l'épreuve adaptative mathématiques

→ l'élève réussit la tâche → l'élève ne réussit pas la tâche

Le contexte proposé est un concours de billes. Il est exposé, dans la première vignette, par un texte relativement court. Nous ne détaillerons pas ici chacune des tâches et renvoyons le lecteur intéressé à l'annexe 8.4 où chaque tâche est détaillée selon une fiche signalétique. Toutefois, nous nous arrêterons ici sur les choix qui ont prévalu à la constitution de tâches complémentaires, présentant quelques éléments de nos analyses *a priori*; il est toujours intéressant de constater, suite aux passations, comment les analyses *a posteriori* compléteront celles-ci (voir chapitre 5) et apporteront des propositions d'amélioration de ces tâches.

La **première tâche complémentaire** (1a) fait l'hypothèse d'une éventuelle mauvaise compréhension des symboles des chiffres utilisés dans la tâche précédente; elle présente ainsi des billes dessinées et non plus symbolisées par un nombre.

La **deuxième tâche complémentaire** (2a) annonce à l'élève qui aurait fait une erreur qu'il n'a pas trouvé les bons sacs et qu'il doit recommencer. D'autres choix auraient pu être faits, par exemple de conserver tous les sacs sélectionnés par l'élève, voire seulement les sacs justes, ou encore d'annoncer le bon nombre de sacs. L'option choisie implique que l'élève recommence sa démarche depuis le début. Nous avons alors prélevé, pour cette tâche complémentaire, l'information qui relève dans quelle mesure l'élève choisit à nouveau les mêmes sacs, révélant peut-être des erreurs qui pourraient être liées à une mauvaise compréhension du nombre.

La **troisième tâche complémentaire** (4a) soumet à nouveau la tâche précédente (relativement classique sur papier et qui consiste à dénombrer des objets, ici des billes, l'élève pouvant les bouger avec la souris) mais en proposant une aide: dix billes ont été mises de côté et ceci est annoncé comme tel à l'élève; toutefois, nous avons fait le choix de ne pas disposer les billes selon une organisation visuellement identifiable (ni alignement, ni repérage possible de sous-groupes) afin que l'impulsion d'organisation du dénombrement soit initiée sans être plus contraignante. Le nombre de manipulations que l'élève effectue ainsi que l'image finale de la disposition des billes sont prélevés, permettant ensuite de repérer si l'élève a investi l'idée d'organiser les billes pour les dénombrer.

La **quatrième tâche complémentaire** (5a) fait suite à une tâche présentant des billes et des boîtes contenant des billes. Cette tâche complémentaire, non seulement réduit le nombre de billes à compter, mais les rend toutes visibles¹⁷. Le nombre choisi est simple, accessible pratiquement à tous les élèves de cet âge (qui plus est promu du cycle précédent), et la représentation laisse voir chaque bille; l'échec à ce genre de question révélerait par conséquent, non pas une difficulté dans l'appréhension du nombre cardinal, mais dans la situation elle-même. La tâche précédente annonce que les boîtes pleines contiennent dix billes et l'élève doit transcrire que trois boîtes correspondent donc à trente billes. En les rendant visibles dans cette tâche complémentaire, la difficulté de se représenter le nombre de billes dans chaque boîte devrait être levée... ou persister.

La **cinquième tâche complémentaire** (6a), réservée aux élèves ayant réussi la tâche précédente, explore la composition non plus seulement en dizaines et unités mais également en introduisant des centaines. Bien que le nombre proposé reste plus petit

¹⁷ Seules les deux « dernières » billes de chaque boîte sont couvertes à moitié par le couvercle; nous avons tenu compte de la possibilité d'une réponse comptabilisant des demi-billes et y reviendrons au chapitre 5.2.3.

que deux cents et donc abordé l'année précédente, il semble intéressant d'explorer la décomposition en centaines, dizaines et unités.

Les analyses doivent ensuite nous permettre de valider ou non ces hypothèses didactiques, ce que nous aborderons dans le point 5.2.3.

4.3.3 Quelques choix dans l'élaboration de l'épreuve papier

Les tâches complexes élaborées pour l'épreuve en mathématiques doivent permettre de dégager dans quelle mesure les élèves investissent leurs connaissances du système décimal avec des nombres inférieurs à deux cents.

Ainsi, la première tâche complexe présente des sièges dans une salle de **spectacle** dont la numérotation n'est que partiellement commencée sur la première et la deuxième rangée. Le nombre recherché est 63 et le répertoire numérique à investir ne devrait pas le dépasser; au maximum, les élèves pourraient-ils travailler dans le répertoire des nombres jusqu'à 100. Inspirée d'une tâche complexe chez nos collègues belges (*Évaluation externe non certificative*. Mathématiques. 2008¹⁸), nous l'avons adaptée en modifiant l'emplacement du siège n° 1 ainsi que la tâche demandée à l'élève (initialement, il ne s'agissait que de compléter la numérotation des trois premières rangées de sièges). Nous avons par ailleurs travaillé graphiquement le dégradé du gris des sièges afin de ne pas induire une démarche souvent systématique des élèves à cet âge consistant à remplir purement et simplement les cases de la suite de nombres commencée. Nous verrons, dans l'analyse des résultats (voir point 5.2.3), que les démarches des élèves demeurent très observables. Elles sont complétées de l'explication demandée aux élèves dans l'encadré réservé à cet effet; consigne était par ailleurs donnée aux enseignants de demander oralement une explication à chaque élève qui remettrait son épreuve sans l'avoir rédigée et d'inscrire eux-mêmes l'explication orale fournie alors par l'élève.

Une autre situation a été développée afin d'observer comment les élèves appréhendent les chiffres¹⁹ composant les nombres. Inspirée notamment d'une activité du moyen d'enseignement romand de 4^e (*Du plus petit au plus grand*²⁰), l'idée de la situation d'une **tombola** a émergé dans le groupe de travail mais présentait l'éventuelle difficulté d'une bonne compréhension et représentation par chaque élève. Les nombres concernés devant être limités à 200, une activité préalable a donc été prévue pour permettre de réactiver le champ numérique concerné et d'assurer une bonne compréhension de la situation aux élèves.

Trois tâches sont soumises aux élèves dans ce cadre :

- la première consiste à poser le principe de billets alignés dans un ordre croissant et à choisir des nombres possibles pour des billets placés entre deux billets donnés (5 billets en tout, 3 billets donnés, 2 billets à proposer; grand nombre de solutions);

¹⁸ *Carnet de l'élève*. Élaboré par le Ministère de la Communauté française. Administration générale de l'enseignement et de la recherche scientifique. Service général du pilotage du système éducatif.

¹⁹ Pour faciliter l'explication et permettre une meilleure compréhension de notre analyse de la tâche, nous distinguerons les chiffres (0, 1, 2, ..., 9) des nombres; ces derniers sont donc composés de chiffres (10, 182, etc).

²⁰ *Fichier de l'élève*. COROME. (1997). CIIP/SR-TI.

- la deuxième reprend le principe de billets alignés dans un ordre croissant et restreint le choix des chiffres à disposition (reste 1, 5, 9) pour créer deux nombres devant encadrer un nombre donné (73);
- la troisième est similaire à la deuxième mais de plus grande complexité, notamment eu égard au nombre de chiffres à disposition (0, 1, 2, 3, 4, 5, 6, 7) et au nombre de billets à trouver; trois nombres sont en effet à construire en-dessous de 89 et un au-dessus.

L'**activité préalable** consiste donc à expliquer aux élèves ce qu'est une tombola et son principe. Contrairement au français, cette activité est proposée environ une semaine avant l'épreuve papier, restant relativement indépendante de cette dernière en termes d'organisation et d'apprentissage. Cette activité est conçue comme un problème posé aux élèves; il s'agit de créer tous les billets de *La tombola*. écrire 200 billets étant très long pour le seul enseignant, les élèves doivent se répartir le travail; une des procédures attendues étant, pour une vingtaine d'élèves, que chacun en écrive 10. Une procédure de contrôle est également demandée afin de s'assurer qu'il n'existe pas plusieurs billets portant le même nombre. Les élèves jouent ensuite quelques fois à *La tombola* en classe.

Cette activité préalable a fait l'objet de nombreuses discussions au sein du groupe. S'il est vrai que, de prime abord, elle semble désigner assez rapidement les connaissances à mobiliser dans les tâches complexes en influençant l'état des connaissances des élèves à ce sujet, elle reste intéressante pour plusieurs raisons :

- la mesure de ces connaissances se fait déjà auparavant à travers l'épreuve adaptative, si bien que l'idée d'influer sur le degré de stabilité de ces connaissances semble alors d'un impact moindre;
- la *mobilisation* (de connaissances) à proprement parler relève de processus cognitifs largement inconnus, dont on parvient à en constater le fait – ou non – sans assurance que la non-mobilisation résulte réellement d'un manque de stabilité de ces connaissances; dans notre cas, réduire quelque peu la part de mobilisation au profit de la combinaison de ressources dans la résolution de tâches complexes n'invalide pas selon nous la mesure de compétence;
- dans le cas d'élèves relativement jeunes, le risque de surcharge cognitive dans la résolution de tâches complexes semble important; leur manque d'expériences de vie rend en effet difficile la mobilisation de ressources dans toute nouvelle situation.

Ces différents arguments ont été pris en compte non seulement dans l'élaboration de l'activité préalable, mais aussi dans celle de l'épreuve adaptative et de l'épreuve papier. La caractérisation de ce que mesure chaque tâche se révélant déterminante pour permettre ensuite une articulation intéressante de ces différentes mesures et informations obtenues.

4.4 Recueil des données

Si le recueil des données est relativement clair pour des épreuves papier, les expériences relatives à des épreuves adaptatives sont relativement récentes. L'essai effectué dans le cadre de ces travaux doit nous permettre de parfaire l'automatisation

du recueil des informations ainsi que leur traitement. Le dispositif élaboré ici a donc été conçu dans une perspective exploratoire, les développements informatisés étant valables pour les tâches spécifiques testées et non comme des « moules » permettant la gestion des variables et la création de tâches d'une même famille²¹.

4.4.1 *Gestion en ligne et récupération automatique des informations*

Les épreuves adaptatives ont été développées pour être utilisables sur des ordinateurs de classes primaires dans les cantons romands, en général des machines non équipées des derniers logiciels et technologies. Trois cas de figure se sont présentés quant à l'équipement des établissements :

- a) deux à trois ordinateurs (disposant d'une connexion internet nécessaire dans notre cas) dans la classe, situation la plus fréquente ;
- b) une salle équipée permettant de faire passer tous les élèves d'une classe en même temps ;
- c) un lot d'ordinateurs portables permettant d'équiper une demi-classe, dans sa propre salle de classe.

Si le cas a) est relativement commun et désormais bien rodé chez la plupart des enseignants, le cas b) pose étonnamment des difficultés supplémentaires relatives à la gestion des questions des élèves simultanément sur les épreuves, mais aussi décalées dans les questions. Le cas c), situation relativement rare actuellement mais *a priori* optimale, permet une passation relativement rapide de tous les élèves et la gestion est facilitée par la disposition de la salle de classe.

Une zone d'administration a également été développée permettant de distribuer les liens auprès des enseignants participant à notre essai et de générer automatiquement, une fois le lien ouvert par un élève, un identifiant et une récupération unique des données. Des protections ont été mises afin d'éviter que des élèves ne quittent accidentellement leur fenêtre de test ou ne répondent pas à certaines questions. Les données ont été récupérées sous forme de tableaux Excel comprenant les informations de toutes les variables pour lesquelles des caractéristiques de récupération avaient été livrées.

Les données sont extrêmement nombreuses et sous des formes diverses nécessitant, dans notre cas et parce que nous ne pouvions bénéficier d'une première expérience, d'un traitement important complémentaire. Un des buts de cet essai est également le tri envisageable des informations pertinentes et exploitables – ou non – afin de dresser un diagnostic des élèves en difficulté. Détaillons ici quelques informations et leurs transformations :

- a) **séquence de manipulations** : en termes d'analyse de réalisation, cette variable est d'une richesse immense et ouvre des perspectives d'analyses didactiques extrêmement vastes. Il a été nécessaire d'analyser les différents cas de manipulations pour comprendre quelles démarches les élèves avaient adoptées, les

²¹ Le terme « famille » est utilisé ici dans une acception relativement large ; les familles de situations évoquées au chapitre 2 sont conçues dans une logique de mesure d'une même compétence, alors qu'ici une situation pourrait, selon le paramétrage des variables, permettre la mesure d'autres compétences.

classer et enfin en définir des catégories de réponses permettant une éventuelle exploitation ;

- b) **temps d'appropriation et de réalisation d'une tâche** : les quelques rares tests préalables avec des enfants de 8-9 ans ne permettant pas de définir des fourchettes de temps idéales pour la réalisation d'une tâche. L'essai effectué ici devait nous permettre de définir dans quelle mesure ces temps sont exploitables ou non ; des catégories ont été définies et analysées pour déterminer leur degré d'information potentiel dans la compréhension des difficultés de certains élèves. Les temps d'appropriation se sont avérés inexploitable (ou très rarement), l'observation en classe révélant des élèves partis en pause ou interrompus pour une information générale de leur enseignant entre deux tâches ;
- c) **retour au texte de référence** : le texte de *Tibili* sur lequel portaient plusieurs questions de l'EA en français étant lu par l'élève au début de la passation, il paraissait important qu'il puisse, selon les questions posées, revenir au texte. Le fait de laisser cette possibilité aux élèves a permis de prélever l'information, souvent inconnue dans ce genre de cas, du nombre de fois où ils y retournent et le temps qu'ils y passent ; une variable a donc été créée (FREL) permettant d'attribuer un code indiquant la réussite potentielle à l'item selon le nombre de relectures ;
- d) **image d'une disposition de points** (tâche mathématique 4 et 4a) : il s'agit d'un cas très particulier de récupération d'informations sous forme d'une image présentant la disposition des billes telle que l'élève l'obtenait à l'issue de sa résolution. Bien qu'il soit *a priori* possible de programmer l'extraction des catégories de réponses selon la disposition de chaque bille, nous avons toutefois besoin d'une première expérience pour vérifier et exploiter les procédures de résolution envisagées. L'exploitation de l'image produite pour chaque élève s'est donc faite par analyse de la disposition globale des billes, avec comme indice pour l'interprétation le nombre de manipulations (voir figures 4.4.1a et b). Si certains élèves procèdent, comme nous le pensions initialement, par rangement en tas de 10 billes (plus ou moins organisés), d'autres placent les billes dans un coin, font des tas non réguliers ou encore, démarche inattendue, accumulent les billes sur la croix. Il a finalement été décidé de ne conserver, pour l'analyse, que la trace d'une organisation par dizaine.



Figures 4.4.1a et b : À gauche, une production d'élève qui procède par alignement de billes par dizaines, formant ainsi quatre lignes et laissant sept billes de côté (52 manipulations) ; à droite une production d'élève qui déplace la plupart des billes dans un coin pour les compter une à une (54 manipulations).

Le nombre important de variables définies et les informations prélevées lors des passations sont, pour un premier essai, essentiels pour définir quelles sont les variables les plus pertinentes à conserver pour l'analyse et, de façon générale, pour les pratiques à venir, si des tests informatisés devaient se développer (comme c'est le cas actuellement dans quelques cantons²²). L'abondance de variables informées permet donc de faire le tri, de concentrer l'information dans des variables reconstruites et menant à des interprétations d'erreurs voire, dans la mesure du possible, à des pistes de remédiation.

Il existe ainsi des variables construites à travers l'ensemble de l'épreuve adaptative, tenant compte de variables simples prélevées initialement ; c'est le cas de la variable FRELCT évoquée plus haut (c) pour l'EA en français et, en mathématiques, des variables suivantes :

- la variable MPRO qui donne un indice, construit sur quatre tâches de l'EA, de la manière dont les élèves entrent dans ces « petits » problèmes ;
- la variable MTT qui fournit un temps moyen sur la réalisation des items obligatoires.

Une variable analysant la navigation dans l'épreuve informatisée a également été construite, en français comme en mathématiques. Elle se base sur plusieurs variables disponibles pour une tâche comme le temps passé sur la tâche, le nombre d'essais ou de manipulation, les tentatives de zapper la tâche²³. Cette variable est déterminante pour créer des profils, comme nous le verrons dans le chapitre traitant de l'analyse des résultats (voir chapitre 5).

4.4.2 Recueil des données des épreuves papier

Les codes préétablis des épreuves écrites ont été soumis aux enseignants qui pouvaient, s'ils le souhaitaient, faire le codage des productions de leurs élèves. Peu s'y sont aventurés si bien que nous ne disposons pratiquement d'aucune remarque de leur part à ce sujet. Toutes les tâches ont été reprises par les groupes disciplinaires respectifs, permettant d'affiner les codes.

Pour le français : les codes ont été établis à partir des tables de spécification, en précisant exactement la signification des critères de réussite, par des exemples en particulier (compréhension de texte). En production, certains critères se sont avérés trop imprécis, sujets à interprétations diverses (notamment les critères relatifs à la syntaxe : qu'est-ce qu'un texte présentant une bonne maîtrise syntaxique en début de 5^e ?). Ces constatations ont été faites durant la correction des textes par les membres du groupe disciplinaire, mais aussi lors de discussions ayant eu lieu par la suite avec plusieurs collaborateurs de l'IRDP. En ce qui concerne la révision du texte par l'élève, nous avons tenu compte des révisions effectives et de leur impact sur une amélioration du texte.

²² C'est le cas du canton de Neuchâtel qui expérimente une partie informatisée dans ses épreuves cantonales au primaire depuis 2010, ainsi que le canton de Genève qui développe également une partie de ses épreuves communes de 11^e sous forme électronique.

²³ Des alertes sont définies pour chaque tâche, indiquant à l'élève qu'il doit répondre et parfois qu'il n'a pas fait l'ensemble de la tâche attendue ; suite à une alerte, certains élèves choisissent de temps à autre une réponse au hasard et passent à la question suivante, mais le temps alors consacré à la tâche est révélateur de la non entrée en matière de certains.

Pour les mathématiques: les observations possibles des procédures élèves ont permis de préciser ou redéfinir les catégories de réponses et codes associés pour certaines variables. C'est notamment le cas de la variable MP1c qui permet de révéler, pour la tâche *Spectacle*, la majorité des démarches utilisées par les élèves. Pour la tâche *La tombola*, la dissociation des variables liées au respect des contraintes (choix des chiffres et ordre des nombres) s'est effectuée durant le codage des réponses.

Dans tous les cas, les choix des variables et catégories de réponses nous semblent mériter deux types d'analyse :

- d'une part, une analyse *a priori* en termes didactiques des productions attendues afin de clarifier les concepts sur lesquels les élèves peuvent buter et, si possible, anticiper par les choix de codes certains types de diagnostics; un prétest s'avère, en ce sens, très riche d'enseignements;
- d'autre part, dans le cas d'un codage par les enseignants (ce qui est souvent le cas dans une épreuve commune externe), une analyse pragmatique de compréhension par les enseignants. Il est en effet nécessaire de s'assurer que le repérage des indices présents dans la réponse d'un élève soit possible grâce à une description claire des critères menant à chaque code pour une variable donnée.

4.5 Avis des enseignants

Revenons à l'expérience de Rey et al. (2004) dont un des apports importants a été l'ambition d'une utilisation directe par les enseignants de leur modèle en trois phases. Dans ce sens, ils ont observé la façon dont des enseignants tentaient de construire des épreuves sur ce modèle. Les investigations de l'équipe de recherche belge se sont ainsi étendues à des actes de formation et ont pu montrer la bonne appropriation du modèle par des enseignants motivés. Qui plus est, ceux parmi les enseignants qui ont testé leur épreuve en classe constatent un diagnostic inédit des élèves et soulignent que la réflexion qu'ils ont menée sur le modèle en trois phases les a conduits à une auto-évaluation féconde de leurs pratiques didactiques et pédagogiques.

L'importance d'une appropriation du modèle d'évaluation par les enseignants eux-mêmes ne fait à nos yeux aucun doute, surtout dans une perspective diagnostique. Le partage du modèle est d'autant plus intéressant qu'un même référentiel, en l'occurrence le PER, vaut pour l'enseignement-apprentissage en Suisse romande. Il nous paraît donc essentiel d'associer dès que possible les enseignants aux réflexions et à la construction des épreuves sur la base du modèle en deux phases.

Déjà, dans la mesure où nous avons pu expérimenter et valider notre modèle en deux phases par le biais d'épreuves, l'une adaptative en ligne et l'autre portant sur une tâche complexe, sur papier, nous avons voulu demander aux enseignants concernés par cet essai leurs avis, une fois avant et une fois après la passation des épreuves dans leur classe.

L'entretien précédant la passation devait permettre une présentation des épreuves par leurs concepteurs et un premier avis des enseignants sur leur manière d'aborder cette passation. La majorité d'entre eux avaient accepté de faire passer les épreuves de français et de mathématiques, d'autres (deux enseignants) seulement celles de mathématiques (soit par intérêt, soit pour des raisons techniques liées à la présence d'un seul

ordinateur dans la classe) et une enseignante celle de français uniquement. Certains ont également précisé quelles étaient les caractéristiques de leur classe à même d'influencer le travail autour des épreuves (élèves allophones, problèmes particuliers en français, accès aux ordinateurs, degré de familiarité avec les tâches complexes, etc.).

L'entretien après passation a conduit les enseignants à faire le point, d'une part sur les aspects positifs et négatifs de l'expérience et du matériel d'évaluation, d'autre part sur ce qui a bien ou moins bien marché durant la passation des épreuves et sur la nature des difficultés rencontrées. Ils ont indiqué également dans quel ordre ils avaient placé les épreuves. Pour terminer, ils ont donné leur avis sur l'utilité diagnostique de celles-ci, et ont formulé des souhaits pour leur amélioration.

4.5.1 *A propos des entretiens pré-passation*

Il est évident que le fait que les enseignants étaient volontaires a favorisé leur soutien à l'entreprise et les a incités à faire le maximum pour que celle-ci se passe au mieux. Peu avant l'entretien, ils avaient reçu et lu l'ensemble des documents relatifs aux deux disciplines. Ils ont donc donné un premier avis sur la base de leur lecture. Dans l'ensemble, ils ont estimé les consignes et les tâches adaptées à leurs élèves. Étant donné que, selon les régions, les classes sont différentes, comportant plus ou moins d'élèves allophones ou en difficulté de français, mais aussi un accès à l'ordinateur plus ou moins facile, certains enseignants ont émis des réserves sur la faisabilité des épreuves pour tel ou tel élève. Ils ont formulé des remarques plus détaillées lorsque l'entretien s'est porté plus spécifiquement sur chacune des disciplines.

En français : si l'épreuve adaptative (EA) a généré quelques appréhensions liées à l'accès et à l'utilisation de l'ordinateur, l'épreuve papier, quant à elle, a occasionné des réserves à propos de son degré de difficulté et de son haut niveau d'exigence. Certains, par exemple, soulignent le fait que l'argumentation n'est pas travaillée avant la 6^e (4P), et que les termes « opinion » (ou « avis ») et « argument » sont difficiles à comprendre pour de jeunes élèves. Par ailleurs, le guide de l'enseignant, très complet, a suscité la crainte d'une longue appropriation par l'enseignant. Des questions de présentation, en particulier à propos du choix des polices de caractère, ont également été évoquées. Enfin, le temps prévu pour l'activité préalable et la réalisation de la tâche complexe a paru court étant donné la densité de cette épreuve, et certains ont estimé nécessaire d'y ajouter une ou deux périodes.

En mathématiques : à nouveau, c'est davantage le contenu de l'épreuve papier qui a amené des remarques que ce qui touche à l'épreuve adaptative. Ainsi, certains ont exprimé leur crainte devant la complexité de la tâche liée à *La tombola*. Plus souvent, ils ont souligné combien les nombreuses consignes et la présence de beaucoup de texte peuvent constituer un obstacle à la réalisation de la tâche de mathématiques. Pourtant, malgré les remarques sur la difficulté des tâches complexes, certains enseignants ont estimé dans le même temps que ces tâches étaient accessibles à leurs élèves.

A priori, les avis ont été favorables, les tâches des deux épreuves (EA et EP) jugées intéressantes et motivantes pour les élèves, aussi bien en français qu'en mathématiques.

4.5.2 A propos des entretiens post-passation

Pour faciliter l'organisation du travail de l'enseignant, ce dernier a eu le libre choix d'établir l'ordre de la passation des épreuves. Or, cet ordre a beaucoup varié d'un enseignant à l'autre, certains commençant par les mathématiques, d'autres par le français, certains par l'épreuve adaptative, d'autres par l'épreuve relative à la tâche complexe. Il ne nous apparaît pas – mais cela reste à vérifier – que cet ordre influence les résultats des élèves. En revanche, il nous semble déterminant, dans une perspective diagnostique et pour rester en cohérence avec les principes du PER, de prendre, une fois les résultats analysés, une position tranchée par rapport à l'ordre des épreuves. Nous y reviendrons plus loin.

Quelques remarques générales: les avis à propos de l'adaptabilité des élèves à l'ordinateur varient d'un enseignant à l'autre. Pour certains, les élèves ont besoin de temps pour se familiariser avec l'appareil et son maniement; pour d'autres, ils n'ont aucun problème à effectuer ces manipulations, se montrant souvent plus habiles que des adultes! Certains ont regretté qu'il ne soit pas possible aux élèves de revenir en arrière dans l'épreuve pour effectuer une éventuelle correction. D'autres ont souligné la différence de rapidité d'exécution des tâches entre les élèves, due en partie à la différence de niveau de lecture existant entre eux.

Il est intéressant de signaler la remarque suivante rapportée par une des enseignantes: elle a indiqué que dans sa classe les élèves avaient rarement l'occasion de réaliser des tâches seuls, sans son aide. En particulier, c'est elle qui initie les démarches de recherche dans des activités de résolution de problèmes. Elle veille également à présenter les activités selon un habillage qui est toujours le même. Elle s'est donc demandé dans quelle mesure elle pouvait laisser les élèves travailler seuls, sans son intervention, dans le cadre des épreuves proposées, et jusqu'à quel point les encadrer si elle décidait de les aider. Le rôle de l'enseignant dans la passation d'épreuves est important et peut influencer les résultats; or, le plus souvent, il est impossible de le contrôler complètement. Une autre enseignante a signalé qu'elle avait pris davantage de temps à réaliser l'activité préalable de français que ce qui était indiqué. Là aussi, il est difficile de savoir ce qui a allongé cette activité et dans quelle mesure elle a fourni des informations supplémentaires qui n'étaient pas suggérées dans le guide. Le fait d'avoir introduit une activité préalable à la tâche complexe, à nos yeux indispensable pour des élèves de ce degré, amène cependant le risque d'un biais dans les résultats à cette tâche, accru par une absence de contrôle des interventions des enseignants, en particulier celles non prévues dans les consignes à leur intention.

À signaler encore que, dans certaines classes, les élèves en grande difficulté n'ont pas pu effectuer l'entier de certaines des épreuves ou, encore, ont dû bénéficier d'une aide individualisée d'un adulte (enseignant titulaire, enseignant spécialisé, chercheur). De tels élèves ont été signalés aux responsables de l'évaluation.

Remarques relatives au français: en ce qui concerne l'EA, un certain nombre de problèmes signalés ont été d'ordre technique; le son ne fonctionnait pas, la souris a connu des ennuis, des pannes ont eu lieu, etc. Pour le reste, les activités ont été appréciées des élèves. Quelques questions ont été problématiques: la segmentation d'un court texte en phrases a été jugée d'autant plus difficile par une enseignante qu'aucun point n'y figurait au départ (ce qui aurait servi d'exemple...). Elle estime que la segmentation de phrases en mots a posé des difficultés de manipulation de la

souris. À propos du texte de compréhension, certains élèves semblent en avoir fait une première lecture rapide pour se lancer très vite dans les questions, avant d'y revenir ensuite. Un enseignant a relevé la grande différence de temps de lecture entre les élèves (donnée que nous avons recueillie et que nous pourrions analyser par la suite). La possibilité de retour au texte (un nombre de fois limité) a été jugée intéressante. De même le fait de pouvoir écouter les consignes a été salué comme positif par certains.

L'EP a posé davantage de problèmes. L'épreuve est dense et longue. Elle a pourtant été généralement appréciée des élèves. Quant aux enseignants, en particulier après la passation, certains ont reconnu la richesse et l'intérêt qu'une telle activité pouvait présenter, pour un travail en classe par exemple. L'avis sur la difficulté de la lettre de Paul varie d'un enseignant à l'autre. En revanche, la majorité estime que ce qui touche à l'argumentation est difficile pour des élèves si jeunes. Des aides, des relances ont été nécessaires et d'ailleurs bien des élèves ont posé des questions sur ce qui leur était demandé de faire. Le questionnaire de compréhension de la lettre de Paul a été jugé très difficile, en particulier lorsque les élèves devaient souligner des passages du texte. « opinion » et « argument » posent des problèmes de compréhension, surtout au moment de passer à l'écriture. Pourtant, un enseignant admet sa surprise en constatant ce que ses élèves ont été finalement capables de produire.

Une majorité observe que les élèves ont de la peine à savoir que corriger dans leur texte, malgré l'aide d'un guide de relecture. Ce sont plutôt les bons élèves qui ont des idées à ce sujet, alors que les autres ne parviennent pas à avoir une vision critique de leur texte.

La remarque suivante a été faite à deux reprises et nous semble intéressante, nous y reviendrons au moment de l'analyse des résultats : la production écrite serait un bon prédicteur du niveau des élèves en français. Si cela devait s'avérer exact, la tâche complexe acquerrait alors une valeur supplémentaire.

Remarques relatives aux mathématiques : l'EA a suscité moins de questions de la part des élèves que celle de français. Ils ont globalement apprécié les activités, même si la difficulté et la longueur de certaines consignes ont été relevées par plusieurs enseignants ; *compter de 10 en 10 en partant de 3* a notamment été jugée difficile.

L'EP en mathématiques a été estimée plus simple et plus facile que celle de français. Pourtant, quelques enseignants ont constaté avec surprise que certains de leurs élèves avaient échoué aux tâches demandées. Par ailleurs, fournir une explication à propos du résultat trouvé est une tâche difficile, même si l'élève avait la possibilité de demander à l'enseignant de transcrire son discours. De plus, le fait que la consigne interdisait l'emploi de la gomme (pour pouvoir suivre les démarches des élèves) a posé problème dans une classe ; certains élèves sont restés bloqués, se privant de revenir sur leur démarche et donc sur leur explication, car ils n'ont pas osé tracer ce qu'ils avaient préalablement écrit.

Quant à la tâche *La tombola*, elle semble plus complexe que celle du *Spectacle* ; quelques enseignants doutent de l'utilité de l'activité préalable pour la résolution des tâches qui ont suivi, certains élèves absents le jour de l'activité préalable ayant par la suite bien compris la tâche *La tombola*. Toutefois, dans une des classes, alors qu'il tentait de se lancer directement dans l'EP sans prendre en considération l'activité préalable, l'enseignant a dû interrompre la passation de l'épreuve pour revenir tout de

même à cette activité tant le besoin d'appropriation de la situation apparaissait nécessaire pour plusieurs élèves. Dans une autre classe, l'activité préalable a été effectuée avec un nombre différent de billets que celui demandé, afin de mieux convenir au nombre d'élèves (220 billets pour 22 élèves); l'enseignante de cette classe relève que ses élèves ont eu de la peine à comprendre ensuite qu'il fallait utiliser des nombres allant jusqu'à 200 pour résoudre les activités de *La tombola*.

Dans une autre classe encore, les consignes ont été jugées trop difficiles pour des élèves allophones, nécessitant une reformulation de la part de l'enseignant. De plus, une amorce s'est avérée nécessaire pour entrer dans la tâche *Le spectacle*; dans ce but, l'enseignant a repris les billets de *La tombola* en les classant selon les nombres pairs et impairs...

4.5.3 *Avis des enseignants sur l'utilité diagnostique des épreuves*

La majorité des enseignants a salué l'utilité de telles épreuves pour leur enseignement. Ainsi, pour l'un d'eux par exemple, elles lui ont permis de travailler la compréhension de la consigne et les relances possibles dans la tâche de production. Il estime en outre que le niveau de difficulté des tâches et des exercices correspond aux compétences de ses élèves. Enfin, ces épreuves sont en adéquation avec sa façon d'enseigner (cet enseignant pratique les séquences didactiques du moyen de français *S'exprimer en français* et n'a donc été surpris ni par la tâche de production ni par le genre de texte portant sur l'argumentation).

D'autres soulignent l'utilité des informations que leur ont fournies les épreuves, sur la numération, sur le fonctionnement logique des élèves par exemple, ou encore sur les textes produits. Cependant, une enseignante pense que pendant la passation de l'épreuve il est difficile de voir ce que font les élèves ce qui remet notamment en cause la feuille d'observation proposée pour l'épreuve papier de français. C'est donc lors de la discussion dans l'entretien qui a suivi la passation qu'elle a pris conscience de la portée des épreuves pour elle et son enseignement. Dans ce sens, elle estime intéressant le fait que les épreuves diagnostiques vont au-delà d'un bilan des compétences des élèves en laissant voir aussi ce que les élèves sont susceptibles d'accomplir ou non par la suite.

Quelques enseignantes auraient bien voulu disposer immédiatement des résultats de leurs élèves aux épreuves adaptatives; attendre est frustrant! De plus, pour tirer bénéfice du diagnostic, elles auraient besoin d'informations fines sur les difficultés rencontrées par certains élèves.

4.5.4 *Améliorations proposées par les enseignants*

Les propositions des enseignants visent essentiellement à l'amélioration des épreuves elles-mêmes. Ainsi, tant pour le français que pour les mathématiques, certains souhaiteraient que la tâche complexe ne comporte pas de question à sous-points, mais qu'à une consigne corresponde une question et une réponse.

Dans l'épreuve informatisée, il conviendrait d'éviter certaines superpositions d'objets sources de difficultés supplémentaires pour les élèves. Certains proposent que d'autres axes thématiques, notamment en mathématiques, puissent également faire l'objet de telles épreuves informatisées. Concernant l'équipement informatique, il est clair que

le fait de ne pas disposer d'un matériel suffisant (en quantité comme en qualité) alourdit la passation, voire la limite.

Globalement, la gestion par l'enseignant de l'épreuve papier, en particulier en français, serait facilitée par une présentation plus structurée du guide pour l'enseignant (présence d'une table des matières). De plus, la majorité est intéressée à disposer, à l'issue des passations, de résultats utilisables et interprétables selon des modes d'analyses prédéterminés et mis en forme à leur intention.

Ces premiers échanges montrent l'intérêt de travailler l'élaboration des épreuves d'évaluation et les suites à leur donner en collaboration avec les enseignants. Ceux-ci sont les mieux placés, en particulier pour dire ce qu'ils attendent d'une évaluation diagnostique qui soit utile à leur enseignement.

4.6 Analyse des consignes

La question des consignes et de leur compréhension par les élèves est tout à fait cruciale et fait l'objet de nombreuses publications. Dans le cadre de son stage à l'IRD, Amélie Guay a fait une synthèse de ces travaux et a entrepris une analyse des consignes de nos épreuves adaptatives de mathématiques et de français (Guay, 2011).

À la suite des observations des différents auteurs pris en considération dans son étude, voici les points de son analyse qui nous semblent intéressants à relever (p. 11-12) :

- les consignes écrites sont toujours plus difficiles à comprendre que les consignes orales ;
- une connaissance lexicale limitée est une source importante d'incompréhension ;
- les relations d'ordre temporel indiquées dans l'énoncé peuvent provoquer des incompréhensions ;
- le sens des verbes porteurs d'une information numérique n'est pas toujours compris par les élèves ;
- le contexte de la consigne est souvent une source de confusion ;
- les consignes longues (en termes de nombre de mots et de nombre de phrases) et qui demandent plusieurs actions différentes sont souvent moins réussies. Elles sont souvent à l'origine des surcharges cognitives des élèves ;
- les phrases complexes (relatives, coordonnées, juxtaposées) et les négations sont les éléments syntaxiques les plus difficiles à comprendre pour les élèves ;
- l'imprécision et la lenteur des processus de décodage chez les lecteurs débutants nuisent à la compréhension des consignes lues ;
- une mise en page défailante peut créer des problèmes de compréhension ;
- une présentation inhabituelle des consignes pour l'élève peut conduire à une mécompréhension de la tâche à mener.

Ces indications sont évidemment de première importance pour des concepteurs d'épreuves, confrontés à la rédaction des consignes. L'analyse de la difficulté des

consignes de nos épreuves adaptatives montre qu'il reste, malgré les efforts déployés pour les éviter, des consignes difficiles à comprendre pour les élèves.

Il ne faudrait toutefois pas penser qu'à des consignes faciles à comprendre correspondent forcément des réussites aux tâches demandées et, inversement, que toute consigne difficile à comprendre entraîne des échecs à la tâche qui lui est associée. Les observations faites par A. Guay montrent en effet que ce n'est pas le cas. En outre, sur la base de critères d'analyse construits par l'auteure, il apparaît que les consignes de l'épreuve en mathématiques sont dans l'ensemble plus complexes que celles de l'épreuve en français. Il y aurait donc là une amélioration à apporter, dans la mesure où, à de jeunes élèves à l'entrée de la 5^e, encore en début d'apprentissage de la lecture, il reste important de présenter des consignes les plus compréhensibles possible.

À la fin de son analyse, A. Guay propose deux outils d'aide à l'analyse de consignes pour les concepteurs d'épreuves: le premier est une série de reformulations de certaines consignes des deux épreuves considérées, le second un guide-conseil pour la rédaction des consignes.

Pour faire comprendre la nature de ce travail, nous présentons quelques exemples d'améliorations de consignes dans chacune des disciplines, en les reliant aux points mis en évidence dans le guide-conseil.

- Le premier point auquel le rédacteur de consignes doit veiller est l'intention de la phrase. Voici quelques propositions de reformulations tenant compte de ce point:

Question 6 de français

Tibili est très étourdi, il a écrit des phrases sans séparer les mots. Place les barres pour séparer les mots à la bonne place.

Tibili a écrit des phrases sans séparer les mots. Place des barres pour séparer tous les mots.

Dans la première phrase, les rédacteurs ont introduit un élément de contexte, mettant en avant l'étourderie de Tibili; or cet élément n'est pas nécessaire pour résoudre la tâche et allonge la consigne.

Question 8 de mathématiques

Paul a reçu pour son anniversaire une boîte contenant cent cases pour ranger ses billes. On peut mettre une bille par case. (...)
--

Paul range ses billes dans une boîte contenant cent cases. Il peut mettre une bille par case. (...)
--

Là encore, comme pour le français, il apparaît inutile de donner une information contextuelle non nécessaire à la résolution de la tâche (anniversaire de Paul).

- La longueur de la consigne est également un point dont il faut tenir compte. Éviter des consignes trop longues doit être un objectif des rédacteurs.

Question 12 de français

Place dans chaque phrase le pronom de conjugaison qui convient. Pioche dans le sac et glisse la bonne étiquette. Attention, tu ne peux utiliser chaque pronom qu'une seule fois.

Place dans chaque phrase le bon pronom de conjugaison.
Attention, toutes les étiquettes du sac doivent être utilisées.

Il est possible de remplacer la relative de la première phrase par un adjectif, plus simple à comprendre. Quant aux deux phrases suivantes, elles peuvent être réduites à une seule, les enfants ayant compris, dès le début du test (par essais de manipulations), que le glissement d'éléments est possible.

Question 4a de mathématiques

Tu as fait une erreur. Lucie veut t'aider : elle a déjà regroupé 10 billes dans un coin. Inscris ici le nombre de toutes les billes lancées :

Tu as fait une erreur. Lucie a regroupé 10 billes dans un coin pour t'aider. Inscris ici le nombre de toutes les billes de Lucie :

Deux propositions séparées par un double point sont ici réduites à une seule phrase, certainement plus compréhensible pour les élèves sous cette forme.

- Le choix des termes utilisés est crucial ; il ne doit pas conduire à proposer des termes trop peu familiers aux élèves.

Question 5 de français

Lis le texte qui suit. À quoi se rapportent les mots encadrés ? ...
Choisis le bon objet ou le bon personnage en cliquant dessus.

Lis le texte. Que remplacent les mots encadrés ? ...
Choisis l'objet ou le personnage en cliquant dessus.

« Se rapporter » est un verbe difficile à comprendre pour des élèves de 5^e ; le verbe « remplacer » semble plus adéquat.

Question 3 de mathématiques

Paul compte de 10 en 10 à partir de 3. Il écrit les nombres obtenus.
Clique sur les nombres ci-dessous qui font partie de sa liste.

Paul compte de 10 en 10 en commençant à 3.
Clique sur les nombres qu'il doit dire.

Nous avons vu plus haut, en examinant les résultats des entretiens avec les enseignants, que cette consigne avait été considérée comme difficile, en particulier sa première partie. Remplacer « à partir de » par « en commençant à » pourrait faciliter la compréhension de la tâche par les élèves. Toutefois, cette tâche étant relativement classique, les élèves associent sa réalisation à une consigne particulière (culture de classe) ; or, si la formulation de celle-ci change, cette association ne se fait pas forcément et les élèves peuvent alors réinterpréter la tâche.

- Enfin, il s'agit encore de veiller à la structure syntaxique de la consigne, par exemple en évitant la présence de subordonnées ou de pronoms tels que « en ».

Question 1 (et 8) de français

Quel est le texte qui résume le mieux l'histoire ?
--

Quel texte résume le mieux l'histoire ?

La consigne corrigée ne contient plus de relative, ce qui l'allège.

Question 8 de mathématiques

(...) Pour ranger ses billes, il remplit d'abord les cases de la première rangée, puis celle de la deuxième et ainsi de suite. (...)
--

(...) Paul commence par remplir la première rangée. Il remplit ensuite la deuxième rangée. (...)

La consigne d'origine, formée d'une seule phrase longue incluant un complément de phrase devient, par la reformulation, deux courtes phrases.

Précisons encore que certains auteurs émettent des réserves quant à une simplification excessive des consignes, dangereuse à long terme dans la mesure où elle empêcherait les élèves de trouver les stratégies leur permettant de faire face aux difficultés de compréhension immédiate. Cependant, si ces réserves nous paraissent légitimes dans le cadre des apprentissages en classe, dans celui de l'évaluation, le risque nous semble grand de perdre en fiabilité de la mesure.

Ces quelques exemples confirment l'importance à accorder à la rédaction des consignes et pourraient servir de point de départ à de nouvelles formulations. Comme la plupart des rédacteurs d'épreuves le constatent eux-mêmes à la suite de la passation, les consignes sont toujours améliorables et les nôtres ne font pas exception : elles mériteraient un travail de reformulation (y compris pour les épreuves papier) ; en ce qui concerne les éventuelles tâches conservées dans une épreuve ultérieure, une comparaison des résultats au regard de l'évolution de la consigne pourrait fournir des indications utiles à la rédaction d'un guide méthodologique d'élaboration de tâches, comme le proposait et l'initiait A. Guay elle-même.

5. Analyse des données des évaluations en deux phases

Nous avons eu la chance, en 2010, de pouvoir procéder, dans le cadre de cette recherche exploratoire, à un essai de passation de nos épreuves dans huit classes de 5^e (3P) réparties dans quatre cantons, ceci au début de l'année scolaire (entre la mi-septembre et la mi-octobre). Qui plus est, le fait de disposer d'une évaluation en français et d'une en mathématiques dans les mêmes délais, conçues dans la même visée et selon le même modèle d'évaluation, nous a permis de proposer aux enseignants intéressés une passation double et nous autorise à exploiter ces informations à la fois parallèlement et en recherchant les liens.

Nos analyses doivent d'abord nous permettre de répondre aux questions suivantes :

- Quelle est la nature et la « gravité » des erreurs repérées ?
- Comment les erreurs sont-elles liées entre elles ?
- Peut-on imaginer fournir une réponse automatisée pour l'EA permettant d'établir des « profils types de difficultés » pour chaque élève²⁴ ?

Il est important de préciser l'enjeu crucial en termes conceptuels autant que pragmatiques de cette dernière question. En effet, de tels « profils » doivent permettre de clarifier auprès des enseignants quelles seraient les difficultés décisives pour la suite des apprentissages des élèves et sur lesquelles ils disposent d'un pouvoir de remédiation ; de plus, ces « profils types de difficultés » doivent conduire à une prise en charge directe et rapide des élèves qui en bénéficient.

Avant d'entrer plus en détail dans le traitement des données obtenues, il nous a paru nécessaire de délimiter les informations qu'il sera possible d'en extraire, ce que nous avons tenté ci-après en interrogeant certaines observations faites au sujet du modèle en deux phases.

5.1 Quelques remarques relatives à notre modèle

Les nombreuses discussions au sein des groupes à propos du choix des tâches complexes et spécifiques ainsi que de leur ordre d'apparition évoquent toute la difficulté de l'ajustement d'une épreuve, difficulté qui existe par ailleurs dans toute création de test. Relevons et clarifions ici certaines remarques critiques, tout en délimitant les buts possibles du traitement de nos données :

²⁴ Le terme « profil » nous satisfait peu étant donné son caractère absolu et alors que nous recherchons précisément l'effet inverse ; cependant, aucun autre terme plus approprié n'a pu être trouvé : il doit donc être compris ici comme un « cliché instantané » de certaines difficultés ; travailler à leur remédiation tend ainsi à faire disparaître ce « profil ».

a) **Les tâches spécifiques de la phase 1 ainsi que l'activité préalable dénaturent la mobilisation des ressources attendues lors des tâches complexes**: s'il est vrai que ces deux étapes préalables offrent l'occasion aux élèves de se remémorer les connaissances qui doivent être mobilisées lors de la réalisation des tâches complexes, il convient également de relever que, sans cela, bien des élèves pourraient se trouver démunis face à la tâche complexe, et que les résultats obtenus alors ne fourniraient que très peu d'informations sur la manière dont les élèves y font face. Or, le fait même que les élèves réinvestissent dans la tâche complexe les connaissances vues précédemment **sans que cela leur soit annoncé** montre qu'ils sont capables d'une certaine mobilisation (certes d'un niveau moindre que ce qu'il est entendu dans les tâches complexes de Rey et al.). Selon nous, de la même façon qu'un adulte reprend un dossier pour s'y replonger avant, par exemple, d'en faire une synthèse, présenter une activité préalable aux élèves leur permet non seulement de repositionner les éléments en jeu (savoir, savoir-faire, ...) mais également de se créer une juste représentation du contexte dans lequel ils seront ensuite interrogés. En procédant ainsi, si la mesure de la mobilisation des ressources est limitée (mesure qui laisse par ailleurs en suspens de nombreuses interrogations), la **combinaison** de celles-ci par les élèves reste intéressante; de plus, l'**appropriation de la situation** nous semble alors être facilitée et l'influence du contexte culturel de chaque élève réduite. Enfin, la part inédite que la situation pourrait présenter est diminuée, ce que nous recherchions (voir Chapitre 2). L'enjeu est donc de produire une information d'abord sur **la façon dont les élèves combinent leurs différentes ressources** et moins sur la façon dont ils les ont choisies.

b) **L'activité préalable permet aux élèves d'effectuer des apprentissages et pas seulement de les réactualiser**. Si effectivement, il semble que certains élèves acquièrent des connaissances à partir de l'activité préalable alors qu'elles auraient dû être intégrées quelques mois auparavant, rappelons que :

- les informations produites dans notre première phase, précisément sur ces ressources, permettent de nous faire préalablement une idée de l'état de ces connaissances chez chaque élève;
- si l'activité préalable permet à des élèves de mobiliser ces ressources dans une tâche complexe ensuite, c'est que soit elles étaient proches d'une certaine stabilisation (degré de compétence oscillant entre 2 et 1), soit la situation leur a permis de faire un saut cognitif sur ces aspects. Dans les deux cas, l'apport est donc bénéfique et le constat d'une différence significative de résultat entre la première et la seconde phase devrait être constatée;
- les élèves testés ici ont entre 8 et 9 ans; d'une part l'expérience à cet âge est restreinte, d'autre part, l'empreinte de différences culturelles ne devrait pas les préteriter dans l'approche d'une situation d'évaluation telle que la tâche complexe l'impose.

Présenter des tâches complexes aux élèves, voire même toute tâche spécifique, provoque, de fait des apprentissages chez les élèves. **Il est impossible de prendre une mesure de compétence de quelque degré que ce soit sans produire des apprentissages chez l'élève, par la réalisation même de tâches complexes, et donc de contribuer à améliorer cette compétence**. Ce trait commun à toute épreuve est d'ailleurs bien connu des enseignants, et il se pourrait que la remédiation proposée consécutivement ne se révèle plus nécessaire pour certains élèves. Mais faire le pari

que les corrections d'une épreuve suffisent à stabiliser les compétences y relatives reste toutefois dangereux, qui plus est déloyal vis-à-vis des élèves, puisque que cela suppose que ceux-ci sont alors les seuls détenteurs du pouvoir d'amélioration de leurs compétences.

L'enjeu est donc bien ici de clarifier quelles sont les possibilités, à partir des données obtenues, de fournir des informations exploitables pour améliorer les apprentissages des élèves, et non pas de classer les élèves selon leurs performances. Nous espérons ainsi produire de la différenciation ciblée selon les besoins des élèves, sans créer de la discrimination; c'est là toute l'importance de la nuance: donner des activités supplémentaires à même de produire les apprentissages attendus chez les élèves en difficulté.

c) Les résultats arrivent souvent trop tard pour être exploitables et ne permettent pas une remédiation suffisamment rapide. Face à certaines difficultés les élèves peuvent alors soit, dans le meilleur des cas, améliorer leur compréhension au contact des apprentissages en cours soit se trouver des stratégies d'accommodation laissant souvent croire à une amélioration de leur compréhension et de leurs connaissances, soit, dans le pire des cas, être bloqués sur les apprentissages consécutifs. L'évaluation en deux phases que nous proposons, notamment grâce à sa partie informatisée, devrait permettre, à l'avenir, de fournir aux enseignants, dès la livraison du matériel de test, les pistes d'analyse qu'ils pourront exploiter. En effet, l'EA, par le traitement automatisé des données tel que nous l'explorons ici, donnerait des pistes de remédiation immédiatement après passation par chaque élève, ces pistes portant d'abord sur les ressources²⁵. L'EP compléterait cette première information afin de fournir des pistes de remédiation ou d'amélioration de pratiques sur les compétences complexes.

5.2 Analyses et résultats

Les analyses effectuées dans le cadre de ce projet permettent essentiellement de proposer des pistes de développement et de réflexion pour les travaux à venir. Le nombre de classes et d'élèves ayant participé à cette passation étant restreint, on ne saurait en tirer des constats généraux sur les élèves. Qui plus est, les enseignants étaient tous volontaires pour se prêter à nos essais, l'implication des élèves en devenant parfaite par rapport à ce que nous pourrions attendre d'une passation à large échelle. Toutefois, les résultats de plus d'une centaine d'élèves issus de quatre cantons différents²⁶ fournissent un ancrage qui nous semble suffisant pour rendre pertinentes ces pistes de recherche dans un processus de développement d'épreuves sur le même modèle, et pour en extraire des propositions de remédiations.

L'enjeu de cette recherche appliquée est aussi d'observer le comportement des élèves face à une épreuve sur ordinateur, pratique encore rare mais qui a tendance à se développer dans quelques cantons: certains cantons en Suisse alémanique utilisent

²⁵ Il n'est toutefois pas exclu de développer certaines tâches complexes sur support informatique; selon le type de traitement des résultats de ces tâches, une partie des résultats automatisés pourrait alors fournir une information sur des compétences complexes (de degré 3).

²⁶ 8 classes ont participé à cette évaluation, 6 ont passé l'évaluation de français et 7 celle de mathématiques. Mentionnons que certains élèves de ces classes présentant des difficultés d'apprentissage aigus demandant un suivi particulier des enseignants n'ont pas pu finir les évaluations et leurs résultats n'ont pas été pris en compte.

entre autres l'outil *Stellwerk*²⁷ destiné à produire des sortes de profils d'élèves en 10^e (8P), notamment dans le but de fournir des indications liées à des profils scolaires et professionnels; Neuchâtel intègre une partie informatisée dans l'épreuve des élèves du primaire en 2012, appréciée globalement des enseignants et des élèves.

La prise en main de l'outil informatique

Nos deux épreuves adaptatives comportent chacune une page destinée à la prise en main par les élèves des fonctionnalités de l'épreuve qu'ils vont devoir passer. Nous y avons également prélevé l'information sur le temps passé et les manipulations effectuées. Les résultats montrent que l'outil est parfaitement maîtrisé par les élèves, même à l'âge de 8 ans.

Par exemple, pour l'EA en mathématiques, dans la partie introductive (voir figure 5.2), la lecture et l'exécution de la première consigne demandant de procéder à un clic sur un objet donné prennent entre 2 et 59 secondes, avec une moyenne de 9,6 sec. et un nombre de clics moyen de 1,2 (tout clic en dehors étant comptabilisé). Pour le principe de cliquer et glisser, le temps varie de 3 à 62 secondes, avec une moyenne de 13 sec. et un nombre de clics moyen de 1,4. Le nombre de clics moyen est un bon indicateur des difficultés de manipulation, non constatées en l'occurrence dans nos EA; le temps est un indicateur moins fiable, étant donné la manière dont certains élèves peuvent parfois être distraits de leur travail en classe par les événements extérieurs. Une variable avait été construite lors des premières analyses pour repérer des élèves ayant des difficultés de manipulation susceptibles de les gêner dans la passation. Or, cette variable a été abandonnée d'une part par manque de fiabilité – les comportements des différentes parties de prise en main étant par trop variables entre eux et la compréhension de la consigne (notamment pour le point c) parfois délicate; d'autre part, du fait de la bonne réussite globale de cette partie et à la suite de l'observation faite des réponses aux tâches pour les rares cas où un doute subsistait.

²⁷ www.stellwerk-check.ch/

Entraînement

Clique sur l'étoile:    

Donne l'ours à la petite fille:
(Clique sur l'ours et glisse-le au bon endroit)     

- Clique sur le 2 et le 4.
- Clique sur le 2. 

Écris le nombre 83 dans la case suivante:
(Clique dans la case avant d'écrire) 

Quand tu es prêt, clique sur la flèche 

Figure 5.2: Page de prise en main de l'outil pour l'EA mathématiques. Chaque partie (a), b), c), d) dans l'ordre) apparaissait au fur et à mesure de la réussite d'exécution de la consigne.

Ces résultats sont peu surprenants au vu des taux de présence et d'utilisation des outils informatiques dans les ménages suisses (voir point 3.3.2).

5.2.1 Méthodologie pour l'analyse de données

Les variables constituées sont de nature essentiellement catégorielle. Bien que nous nous soyons efforcés de construire des codes destinés à repérer des erreurs types, de la réponse juste à celle « peu grave » puis « un peu plus grave »²⁸, ces codes ne sont pas comparables dans leur valeur mais bien dans l'erreur repérée. Rappelons que nous avons orienté nos évaluations dans une visée diagnostique, si bien que nous avons constitué parfois des catégories qui sont finalement investies par un nombre relativement faible d'élèves. Analyser des réponses à une épreuve diagnostique de compétences qui permette de repérer des difficultés éparses et diverses, loin d'une répartition homogène: là est toute la difficulté.

La démarche statistique que nous engageons est d'abord exploratoire et une grande part de nos analyses relèvent donc de statistiques descriptives dans le but de formuler un certain nombre d'hypothèses relatives aux erreurs des élèves. Par conséquent, nous ne nous engageons que partiellement dans des statistiques inférentielles, les statistiques descriptives devant permettre de privilégier certaines pistes de traitement des résultats pour les travaux futurs.

²⁸ En règle générale, nous avons retenu le code 1 pour la réponse juste, puis 2 pour une erreur peu grave, 3 une erreur un peu plus grave, etc. Mais l'écart entre deux codes ne porte aucune distance régulière entre deux erreurs. Le code 0 est réservé à une réponse erronée dans le cas de variables ne repérant que la réussite, le code 9 quant à lui représente la non-réponse.

Actuellement, la tendance pour la plupart des évaluations de compétences est d'utiliser une méthodologie IRT, souvent très contraignante pour les items et leur codage. Selon notre connaissance préalable des contraintes d'application d'un IRT, étayée par des avis de spécialistes tels que Bain (2010), les outils IRT ne semblent pas pertinents quand on espère précisément valider l'acquisition de connaissances et de compétences individuelles. En effet, l'utilisation de ces outils tend (...) à *transformer un test d'acquisition en test d'aptitude, à passer d'un objectif sommatif à un but pronostic, sans que les critères de validité soient ajustés et vérifiés en conséquence* (p. 57). Il nous a alors paru nécessaire de nous tourner vers d'autres types d'analyses, d'explorer celles les plus à même de nous fournir les informations recherchées et de répondre à nos questions.

Nous ne nous sommes donc pas donné de contraintes quant à la discrimination des tâches, considérant au contraire qu'il était utile de connaître les quelques élèves ayant encore des difficultés sur des tâches relativement simples. Nous ne nous sommes pas non plus forcés à construire, pour chaque variable, des catégories de réponses qui disposeraient d'une répartition homogène. Les épreuves doivent rester abordables par tous les élèves, la majorité devant les réussir. Nous tentons donc d'extraire une information relativement subtile sur des erreurs que nous espérons repérables selon les connaissances didactiques actuelles. Nous verrons qu'étendre une épreuve à la réalisation de tâches complexes mène au repérage de certaines difficultés peu explorées jusqu'à présent.

Construction de « variables transversales » aux tâches

La possibilité de regrouper des informations pour créer de nouvelles variables, comme celle caractérisant la prise en main de l'outil informatique²⁹ a déjà été évoquée plus haut. À l'instar de ce que Roegiers (2004) propose, nous avons tenté de construire des « variables transversales » relevant des critères relatifs à la gestion et la réalisation de tâches complexes.

En français, trois critères ont été retenus, relatifs à l'EP :

- appréhension du contexte ;
- traitement du contenu ;
- textualisation.

Ces critères ont été décrits plus haut au point 4.3.2 et nous ne revenons pas ici sur leur définition. Ils sont issus des tables de spécifications présentées, et rendent compte d'activités langagières propres à tout travail sur le texte, en lien avec le genre de texte envisagé.

En mathématiques, nous avons tenté de créer des variables communes d'une part à l'EA et d'autre part à l'EP :

- entrée dans le problème, basée sur des variables retraçant l'entrée dans la situation pour quatre tâches de l'EA ;

²⁹ Cette variable n'a finalement pas été retenue dans la mesure où elle n'apportait pas assez d'informations exploitables pour un diagnostic.

- temps moyen de réalisation des tâches obligatoires de l'EA;
- type de démarche mise en œuvre (experte, exploratoire, ...), basée sur l'analyse de résolution des tâches complexes dans l'EP.

Nous avons également tenté de déterminer des variables similaires pour le français et les mathématiques dans le but d'interroger la transposabilité des démarches de résolution de tâches complexes. Il s'agit de variables que décrivent les critères suivants :

- pertinence : l'élève entre correctement dans la tâche, choix des outils, bien compris;
- utilisation correcte des outils;
- lien entre la situation et la réponse.

Nous reviendrons sur les variables qui ont permis leur construction (voir point 5.2.4), mais relevons que si les informations à disposition permettent de les construire avec assez de satisfaction, il serait intéressant, à l'avenir, de prendre en compte cette volonté dès la conception des épreuves. Ces trois critères pourraient notamment être utilisés sur trois tâches complexes dans une même discipline afin de voir dans quelle mesure un élève présente régulièrement des difficultés sur l'un d'eux. Roegiers (2004) propose par exemple que si un même critère est réussi pour deux tâches parmi les trois, on peut admettre que la compétence correspondante est suffisamment bien développée, ce qu'il nomme la règle des 2/3 (p. 188).

Analyse en composantes multiples

L'une des analyses statistiques utilisée dans la présente recherche est l'analyse des correspondances multiples (ACM). Elle (...) est une *technique de description de données qualitatives* : on considère ici n individus décrits par p variables. (...) Cette méthode est particulièrement bien adaptée à l'exploration d'enquêtes où les questions sont à réponses multiples. (Saporta, 1990, p. 217). Il s'agit d'une méthode statistique qui a pour but de s'intéresser à plusieurs variables qualitatives, catégorielles ou nominales. Plus concrètement, l'ACM est une analyse factorielle dont l'intérêt premier est de pouvoir homogénéiser les données et mettre en lumière des liens non linéaires entre les variables, leurs modalités, ainsi qu'entre les individus. La proximité, en projection sur le plan factoriel, des points moyens des catégories représente plusieurs individus.

L'interprétation d'une ACM se situe au niveau des oppositions entre les différentes modalités ou groupes d'individus ainsi qu'au niveau des proximités entre ces derniers (voir figure 5.2.1).

Dès le moment où deux modalités d'une même variable sont proches, les individus possédant l'une de ces modalités et ceux qui possèdent l'autre modalité sont considérés comme similaires au niveau des autres variables. L'interprétation qui pourra en être faite est l'existence de profils semblables. De la même manière, si deux modalités d'une variable différente sont proches, on pourrait analyser le lien éventuel entre ces modalités données pour des variables différentes, menant à des hypothèses didactiques sur les sources de certaines erreurs (voir point 5.2.3 Analyse EA Math).

Il est important de préciser ici que les axes factoriels se déterminent grâce aux variables dites actives (ou principales) et permettent l'interprétation des proximités et distances

des catégories des variables principales ou des variables dites supplémentaires (parfois intitulées de « type erreur »³⁰). La première interprétation peut concerner la recherche des variables actives (principales) dont la contribution cumulée est supérieure à 1/p. Elles déterminent les axes tandis que les variables supplémentaires ne participent pas à cette détermination. Une deuxième interprétation, plus détaillée, permet de rechercher les modalités dont la contribution est supérieure.

Il peut sembler étonnant, eu égard à certaines de nos catégories de réponses faiblement dotées, de choisir l'option d'une ACM, sachant que ces catégories se retrouvent souvent loin des autres catégories sur le plan factoriel. En l'occurrence, c'est moins le traitement de ces catégories qui nous intéresse à travers l'ACM – bien que cela donne parfois quelques pistes – que celui des catégories plus fortement représentées et qui doivent permettre d'observer si l'expression d'une compétence comporte des traits communs.

De plus, nous avons recouru à une analyse en clusters dont les modalités sont les suivantes. Deux individus sont proches s'ils ont choisi les mêmes modalités. Ils sont éloignés s'ils n'ont pas répondu de la même manière. On suppose au départ que l'ensemble des individus est muni d'une « distance » calculée à partir des points-individus. On vise à diviser l'ensemble d'individus en différents « groupes » homogènes (*clusters*), en ce sens que les données de chaque groupe partagent des caractéristiques communes, qui correspondent le plus souvent à des critères de proximité que l'on définit en introduisant des mesures et classes de distance entre objets.

Il est à préciser ici que les premières analyses en composantes multiples que nous avons effectuées nous ont permis d'affiner nos variables et nos catégories. D'une part nous avons pris des décisions quant aux variables principales (actives) en combinant parfois plusieurs données, d'autre part, nous avons décidé des variables secondaires à conserver et à projeter sur le plan factoriel. Notons que nous avons d'abord construit nos variables principales en les considérant comme porteuses d'une information relative au type de réalisation de la tâche, avant de définir la réussite à la tâche ou à l'épreuve.

Nous ne présenterons pas ici ces premiers affinages, mais nous nous concentrerons sur les analyses qui ont été menées dans un deuxième temps, lorsqu'elles nous semblent porteuses d'hypothèses intéressantes. Mentionnons au passage que, lors de ces premières analyses, nous avons pu constater une répartition des variables relatives au canton, au genre ou à la langue maternelle correspondant étonnamment bien aux constats d'autres recherches comme PISA par exemple, tendant à considérer notre population à disposition comme assez représentative somme toute du paysage romand.

5.2.2 Les résultats des épreuves de français – cycle 1

L'ÉPREUVE ADAPTATIVE (EA)

Cette épreuve contenait 12 questions touchant à la compréhension d'un texte (*Tibili le petit garçon qui ne voulait pas aller à l'école*) et au fonctionnement de la langue, dont

³⁰ Nous éviterons d'employer ce terme, induisant le doute entre nos catégories permettant de repérer des erreurs dans les productions des élèves, catégories existant également pour nos variables principales.

deux appelaient, en cas d'échec, une question complémentaire. La première question était reposée à ceux qui y avaient échoué, au milieu du test cette fois et ce, dans le but de voir si la tâche était mieux réussie grâce aux questions précédentes et à d'éventuels retours au texte. Par ailleurs, un grand nombre de variables globales, permettant une prise d'informations sur les caractéristiques des élèves, les conditions de la passation (temps de lecture du texte, retour au texte, recours à l'écoute des consignes, etc.) ont été prises en considération.

De cet ensemble impressionnant de données, nous avons réalisé un tri pour ne retenir finalement que les informations pouvant nous fournir des résultats exploitables à court terme, c'est-à-dire celles susceptibles d'entrer en jeu dans le diagnostic que nous recherchions. Si cela s'avérait utile, il serait toujours possible, dans des études ultérieures, de reprendre tout ou partie de cet ensemble d'informations.

Variables retenues

Variables portant sur les caractéristiques de l'élève: le genre, l'âge et la langue parlée à la maison des élèves.

Variables globales:

- le temps de lecture du texte de *Tibili le petit garçon qui ne voulait pas aller à l'école*;
- le temps moyen par item obligatoire (items soumis à tous les élèves);
- l'efficacité du retour au texte lorsque la question le permettait;
- le nombre de fois où l'élève semble avoir passé trop vite d'un item à un autre;
- la réussite globale à l'épreuve (sans tenir compte des questions complémentaires).

Variables-réponses

A une exception, toutes les questions, obligatoires et complémentaires, ont été retenues dans un premier temps et leurs résultats calculés. La question F11, demandant de trier des mots de la même famille, a posé un problème de fond qui nous a décidés à l'écartier. Une analyse *a posteriori* a mis en évidence un choix inapproprié de mots pour chacune des familles à reconstituer, et a donc remis en cause le contenu de la question. Finalement ce constat a conduit à ne pas tenir compte des résultats de cette question, dans la mesure où cette dernière n'apportait, en l'état, aucune information intéressante pour le diagnostic. Relevons ici qu'il ne s'agit pas de mettre de côté des tâches très bien réussies par les élèves; d'excellents résultats attestent en effet que les élèves ont bien intégré les connaissances qui sous-tendent les tâches correspondantes.

Premiers résultats descriptifs

Dans un second temps, nous voulons présenter les résultats en termes de pourcentages à chacune des catégories de réponses (voir les tâches dans l'annexe 8.3).

Tableau 5.2.2: Résultats par question en pourcents

%	F01	F01a	F02	F03	F04*	F05	F06a	F06b	F06c
Réussite/ réussite partielle	78	46	79	60	1: 41 2: 22	1: 16 2: 45	1: 63 2: 28	1: 75 2: 17	1: 72 2: 17
Réussite minimale/ non réussite	22	54	21	40	3: 24 4: 13	3: 39	3: 9	3: 8	3: 11

Pour F04: code 1 = tout juste; code 2 = deux réponses justes; code 3 = une réponse juste; code 4 = aucune réponse juste.

Pour F05: code 1 = juste; code 2 = 1^{ère} phrase juste puis au moins une faute; code 3 = autres erreurs.

Pour F06: code 1 = tout juste; code 2 = n'a fait que des oublis; code 3 = présence de non-mots (autres fautes que oublis).

%	F07	F08	F09	F09a	F10	F12
Réussite/ réussite partielle	1: 42 2: 35	1: 28	1: 50 2: 24	1: 21 2: 34	1: 12 2: 6 3: 33	1: 53 2: 12 3: 19
Réussite minimale/ non réussite	3: 23	2: 44 3: 28	3: 26	3: 45	4: 4 5: 32 6: 12	4: 16

Pour F07, F09 et F09a: code 1 = tout juste; code 2 = une seule erreur; code 3 = plus d'une erreur.

Pour F08: code 1 = réponse juste; code 2 = réponse fausse identique à celle de la question 1; code 3 = réponse fausse différente de celle de la question 1.

Pour F10: code 1 = tout juste; code 2 = tout juste mais incomplet; code 3 = erreurs plausibles; code 4 = points seulement en fin de ligne; code 5 = autres erreurs avec 5 points ou moins; code 6 = autres erreurs avec 6 points ou plus.

Pour F12: code 1 = tout juste; code 2 = erreurs à consonnance correcte; code 3 = une seule erreur à consonnance non correspondante; code 4 = autres erreurs.

Plusieurs questions, à plus de 70% de réussite³¹, mettent en évidence une bonne performance des élèves, étant admis, pour certaines d'entre elles, une erreur possible: F01, F02, F06, F07, F09.

En ce qui concerne les questions complémentaires, il est intéressant de constater que 28% des 22% d'élèves qui ont échoué F01 parviennent à se corriger lorsque la question leur est présentée à nouveau dans F08, soit 7 élèves sur les 25 qui ont échoué à F01; 44% de ces 22% redonnent la même réponse fausse qu'à F01 (soit 11 élèves), alors que 28% proposent à F08 une autre réponse qu'à F01, mais toujours fausse (soit 7 élèves)³².

La moitié des élèves (57 sur 115) ont fait au moins une erreur à F09 et ont donc passé la question complémentaire F09a. La moitié d'entre eux réussissent la question complémentaire ou ne font qu'une seule erreur. Reste une moitié qui éprouve encore

³¹ Le choix de 70% pour parler de réussite à une question est arbitraire; c'est la limite que nous avons fixée, sachant que, dans une perspective diagnostique, l'important se situe davantage dans la mise en évidence des principales difficultés rencontrées par les élèves que dans les taux de réussite.

³² Il est évident qu'au vu des petits effectifs d'élèves selon les catégories d'erreurs, il s'agit de rester prudent sur la signification des pourcentages de résultats.

des difficultés à la reconnaissance de sons. À signaler tout de même, et pour nuancer ces constats, que le matériel à disposition dans certaines classes n'a pas favorisé une écoute optimale des mots proposés dans F09a. Il faudrait donc d'autres investigations pour corroborer nos résultats.

Analyse des correspondances multiples

En nous penchant sur les graphiques issus de cette analyse, nous avons constaté la difficulté d'interpréter à la fois la signification des axes et la nature de la proximité des questions dans les espaces établis par ces axes.

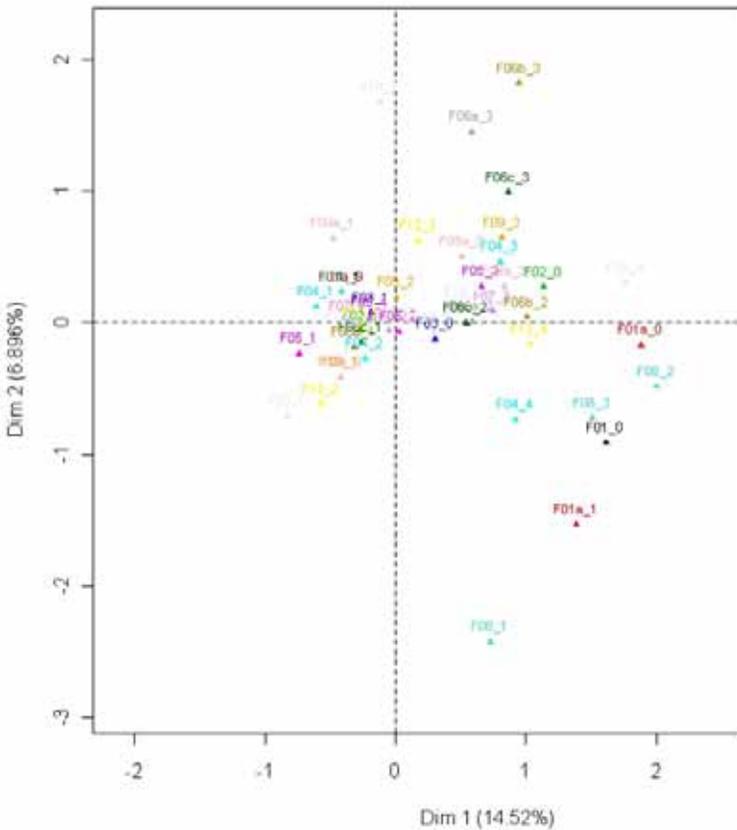


Figure 5.2.2a : Répartition des catégories de variables principales pour l'EA français

La plupart des différentes modalités de réponses aux questions se répartissent dans les espaces créés par les axes, même si les modalités de réussite totale sont davantage regroupées sur la gauche de l'axe horizontal que celles liées aux mauvaises réponses, sur la droite. Ainsi, cet axe, expliquant environ 15% de la variance, signifierait réussite-échec, les bons résultats se lisant sur la gauche du point de rencontre des axes, les moins bons sur la droite, quant à eux plus largement répartis dans la hauteur. En revanche, le deuxième axe reste très difficile à identifier. Nous avons encore relevé

la place particulière tenue par la modalité non réussite de la question F01 (F01_0), occupant, avec ses questions complémentaires, F01a et F08 (les modalités d'échec), l'espace à droite de l'axe horizontal et dans la partie inférieure de l'axe vertical.

Discussion des résultats et hypothèses

À la suite de ces différents constats, nous formulons un certain nombre de remarques et d'hypothèses.

Quelques tâches s'avèrent difficiles pour les élèves de cet âge : F03, dont on peut dire qu'elle dépasse ce qu'on peut attendre de la fin du cycle 1, F04, F05 et F10, également trop exigeantes pour cette fin de cycle 1 :

- à propos de F03, il est en effet trop tôt pour que des élèves de fin de cycle 1 soient à même de reconnaître le genre d'un texte, surtout si, en classe, ils n'ont pas eu l'occasion de travailler la reconnaissance de certains genres et les indices permettant de les identifier ; relevons tout de même que plus de la moitié des élèves ont répondu correctement à cette question ;
- F04 demande aux élèves d'associer des répliques de dialogue aux personnages correspondants. Pour cela, ils doivent à la fois faire appel à leur mémoire (mais ils peuvent aussi revenir au texte), s'appuyer sur des indices du texte pour retrouver le personnage qui parle, avoir compris le sens global du conte. Au vu des résultats, ce seraient donc les bons lecteurs qui parviennent à réussir complètement cette question ;
- F05 présente successivement, à l'intérieur d'un morceau de texte issu du conte, trois phrases comprenant un pronom entouré ; il s'agit pour les élèves d'indiquer, parmi quatre propositions, celle que remplace ce pronom. Si la première phrase, relativement simple (« il » pour Tibili), a été plutôt réussie, les deux suivantes, plus complexes (« l' » pour le coffre et « lui » pour Tibili) ont donné lieu à un grand nombre d'erreurs. Là encore, nous émettons l'hypothèse que seuls de très bons lecteurs peuvent parvenir à une réussite totale de cette question ;
- F10 est particulièrement difficile dans la mesure où les élèves doivent connaître la signification de la phrase écrite et des indices qui la fondent, ce qui ne peut être le cas en fin de cycle 1. Face à un texte sans majuscule ni ponctuation, ce qui représente une situation rarissime aussi bien dans la réalité sociale qu'à l'école, les élèves devaient placer les points marquant les différentes phrases d'un court texte. La tâche s'est en effet révélée complexe et très peu l'ont réussie entièrement. Ceux qui y sont parvenus ont fait preuve d'une connaissance grammaticale de haut niveau pour leur âge.

À propos de l'ACM, en tenant compte de certains regroupements de questions en fonction des axes, nous avançons les points suivants :

- indépendamment de la réussite ou non à la question, la répartition des questions met en évidence deux regroupements de réponses, d'une part un regroupement lié au fonctionnement de la langue (en particulier F06 et F09, mais aussi F10 et F12, généralement plus difficiles), d'autre part un autre lié à la compréhension de texte (il s'agit de F01, F02, F04, F05, F07) ;

- la non-réussite à la question F01, associée à la non-réussite à F04, pourrait laisser apparaître un type de difficulté particulier en lien avec la compréhension du texte. Nous nous demandons si le fait de ne pas réussir à identifier ce qui résume le mieux une histoire pourrait être révélateur de difficultés majeures en compréhension de texte. Il s'agirait donc de regarder plus attentivement les résultats des élèves ayant échoué à la question F01.

Clusters d'élèves et types de difficulté

L'analyse des correspondances multiples a conduit à une analyse en *clusters* permettant de répartir les élèves dans des groupes homogènes. Si ces *clusters* départagent en fait les élèves des meilleurs aux moins bons, il peut être intéressant, au-delà de ce type de constat, et en fonction de nos hypothèses, de tenter de comprendre la nature des difficultés rencontrées par les plus faibles d'entre eux, se souvenant que nous sommes dans une évaluation diagnostique qui, précisément, devrait guider l'enseignant dans son travail pour remédier aux difficultés les plus patentées encore rencontrées par certains de ses élèves.

Trois *clusters* ont été déterminés. Si le premier rassemble effectivement les élèves ayant un fort taux de réussite à l'ensemble des questions, les deux autres révèlent deux types d'échec à ce test. Il faut encore préciser que la taille des *clusters* est très différente de l'un à l'autre, puisque le premier comprend 86 élèves, le second 7 et le troisième 15 élèves. Le profil du deuxième *cluster* est difficile à identifier et nous le laissons pour l'heure de côté. La principale caractéristique du troisième *cluster* est un échec marqué aux questions F06 (a, b et c) et F09, portant l'une sur la segmentation de phrases en mots, l'autre sur la reconnaissance graphème-phonème. Or, ces questions sont déterminantes pour l'apprentissage de la lecture et les attentes fondamentales qui les sous-tendent devraient être atteintes en fin de cycle 1 pour permettre les apprentissages dans le cycle suivant. Il y aurait donc là à approfondir les résultats de ce groupe pour mieux appréhender ce qui le caractérise dans le cadre de ce test. Dans ce sens, nous sommes allés vérifier qui étaient ces élèves et en avons retenu 11 qui nous paraissaient en effet relever de ce *cluster*. Ces élèves n'ont guère de difficulté à répondre aux questions portant plus spécifiquement sur la compréhension de texte (F01, F02 et, dans une proportion équivalente aux meilleurs élèves, F04 et F07). F03 (sur la reconnaissance du genre de texte, dont l'exigence dépasse celle du cycle 1) et F05, relative à l'identification d'anaphores dans un court texte, leur posent des problèmes. Ces constats semblent donc aller dans le sens d'un spectre de difficultés tournant autour du fonctionnement de la langue, ou en tout cas, des attentes fondamentales de ce domaine, centrales pour les premiers apprentissages de la lecture.

Un type de difficulté particulier à F01

Ensuite, dans le droit fil de nos hypothèses et pour donner suite à cette analyse en *clusters*, nous avons examiné plus en détail les résultats des élèves qui avaient échoué à la question F01, en nous demandant s'ils pouvaient provenir d'un type de difficultés particulier. Notre démarche est ici qualitative, d'observation des données à disposition; elle n'est pas consolidée par des vérifications statistiques. Parvenus à la fin de notre mandat, il ne nous a en effet pas été possible d'aller au-delà de ces explorations et des questions qu'elles soulèvent. Toutefois, l'accès aux données restera possible si

de nouvelles investigations devaient être mises sur pied pour consolider nos résultats et parfaire l'évaluation diagnostique en Suisse romande.

Nous avons dénombré 26 élèves qui avaient échoué à la question F01. Parmi eux, 6 ont ensuite réussi la même question placée plus loin dans le test. Il reste 18 élèves qui peuvent être considérés comme ayant véritablement échoué à F01 et qui présentent une suite de réponses aux caractéristiques propres. En voici une brève description :

Ce groupe est composé d'une majorité de garçons, dont la moitié a tendance à lire le texte de départ trop vite (N=9 élèves) (une minute ou moins) et quelques-uns (N=4) très lentement (plus de 5 minutes). Beaucoup vont relire le texte de *Tibili le petit garçon qui ne voulait pas aller à l'école* à l'occasion d'une question ou d'une autre, mais leur relecture est peu productive et conduit le plus souvent à un échec à la question. Une petite moitié environ (N=7) de ces élèves ne réussissent que 0 à 4 items obligatoires de l'épreuve adaptative, alors que tous les autres en réussissent de 5 à 7.

La majorité de ces élèves échouent aux questions portant sur la compréhension globale du texte, sur la reconnaissance de dialogues en fonction de certains personnages, sur l'identification des anaphores présentes dans un extrait de texte. En lien avec le fonctionnement de la langue, ils éprouvent de la difficulté à identifier des phrases écrites d'un court texte (mais la tâche dépasse les attentes du cycle 1) et à reconnaître certains phonèmes ainsi qu'à associer certains phonèmes-graphèmes.

Que retenir des résultats de EA ?

- Un peu moins de la moitié des élèves réussit bien l'épreuve adaptative, c'est-à-dire entre dans les catégories 1 et 2 du critère général de réussite (FREU 1 ou 2)³³ ;
- il nous est apparu qu'un petit nombre d'élèves (N=7) échouent l'ensemble de l'épreuve adaptative, aussi bien les questions de compréhension que celles sur le fonctionnement de la langue. Nous regarderons plus loin ce qu'il en est pour eux (mais aussi pour les élèves des autres groupes) de leurs résultats dans l'épreuve papier à propos de la tâche complexe ;
- finalement, il existe grosso modo trois types de difficultés à cette épreuve : une touchant plus particulièrement la compréhension de texte, les questions portant sur le fonctionnement de la langue étant par ailleurs majoritairement réussies ; une autre, au contraire, centrée essentiellement sur les questions de segmentation de phrases en mots et de reconnaissance phonème-graphème, relatives à des attentes fondamentales de fin de cycle 1, alors que la plupart des questions de compréhension sont réussies ; enfin, pour un très petit groupe d'élèves, une difficulté générale entraînant un échec sur l'ensemble des questions de l'épreuve ;
- la vitesse de lecture du texte de départ joue un rôle sur les résultats ; les élèves connaissant des difficultés dans l'épreuve ont tendance à lire le texte soit très vite (trop vite pour qu'il leur soit possible de répondre aux questions qui suivent) soit très lentement (ce qui pourrait signifier encore une difficulté à déchiffrer). De même, le retour au texte, pendant le test, peut mener à la réponse exacte ou non.

³³ FREU 1 signifie que l'élève a réussi entre 11 et 13 items obligatoires ; FREU2 qu'il a réussi entre 8 et 10 items obligatoires ; FREU 3 qu'il a réussi entre 5 et 7 items obligatoires ; FREU 4 qu'il a réussi entre 0 et 4 items obligatoires. La répartition en pourcents se présente de la manière suivante : FREU 1 : 7.3% ; FREU 2 : 37.6% ; FREU 3 : 35% ; FREU 4 : 16.2%.

Ce dernier cas de figure est bien sûr plus fréquent chez les élèves en difficulté. Ces derniers ne semblent ainsi pas tirer profit de ce retour, peut-être ne sachant pas comment exploiter le texte qui revient sur l'écran pour y trouver l'information cherchée.

L'ÉPREUVE PAPIER (EP)

Le groupe de travail de français a effectué lui-même le codage et la correction des épreuves papier. Dans certains cas, la correction a conduit les membres du groupe à modifier ou à préciser le contenu de plusieurs critères. Il faut bien l'admettre : corriger une tâche complexe n'est pas aisé. D'ailleurs l'équipe belge (Rey, Carette, Defrance & Kahn, 2004) avait expérimenté cette difficulté et avait finalement construit une échelle graduée de la réussite de la tâche (interdisciplinaire), sans détailler finement ce que représentait cette réussite. Pour notre part, nous avons choisi une autre option et les tables de spécification présentées au point 4.2.3 dévoilent les critères retenus, répartis dans les composantes de la tâche, qu'elle soit de compréhension ou de production.

Variables retenues

Les *variables portant sur les caractéristiques de l'élève* sont le genre, l'âge et la langue parlée à la maison des élèves.

Les *variables globales* concernent la production écrite uniquement et portent sur l'ajout d'un contenu d'une part, la modification du texte au niveau de la forme langagière d'autre part.

- En ce qui concerne l'ajout ou non d'un contenu supplémentaire, non demandé par la consigne de production, il pourrait être le signe soit d'une mauvaise compréhension de cette consigne, soit de l'aisance du scripteur à écrire une lettre qu'il étoffe d'informations complémentaires.
- Toute modification apportée au niveau de la forme langagière du texte lors de la relecture (à l'aide d'un guide de relecture du texte), au lendemain de la production de texte, pouvait contribuer à améliorer la production (code 1) ou donner lieu à au moins une erreur dans le texte (code 2).

Variables-réponses

En compréhension : Les questions ont été retenues et les codes ont, dans la plupart des cas, ont été réduits (deux codes, par exemple, fondus en un seul, en particulier pour les questions demandant un soulignement, plusieurs d'entre eux étant regroupés soit sous *soulignement attendu* soit sous *soulignement partiel*, soit sous *autre soulignement*).

En production : Les critères FP1 et FP5, relatifs à la maîtrise du format de la lettre à écrire (en-tête et signature) ont été regroupés en un seul, FP1.

FP6, portant sur la syntaxe et donnant lieu au départ à différents sous-critères, a posé problème. En particulier, il s'est avéré très difficile de mettre d'accord les correcteurs sur ce qui fait qu'un texte répond au critère d'un texte écrit et ne contient pas de formes propres à l'oral. Ce critère a donc été supprimé. Seules les traces d'une non-

segmentation en mots et/ou d'un ordre inadéquat des mots dans la phrase ont été finalement relevés, au sein d'un seul critère regroupant ces deux sous-parties.

Premiers résultats - En compréhension

Tableau 5.2.2e: Résultats par question en pourcents

%	FQ1	FQ2	FQ3	FQ4	FQ5	FQ6	FQ7a	FQ7b
Réussite/réussite partielle	1: 98	1: 81	1: 38 2: 2	1: 38 2: 4	1: 81	1: 76	1: 68	1: 36
Non réussite	0: 2	0: 19	0: 58	0: 48	0: 16	0: 13	2*: 21 0*: 1	0: 24
Non réponse	9: 0	9: 1	9: 3	9: 10	9: 4	9: 11	9: 10	9: 40

1 : réponse ou soulignement attendu – 2 : soulignement partiel.

0 : autre réponse, autre soulignement.

9 : aucune réponse ou aucun soulignement.

2* : réponse erronée.

0* : plusieurs cases sont cochées (une juste et une fausse).

Cinq questions sur les huit donnent lieu à de bons, voire de très bons résultats :

- FQ1 demande aux élèves qui a écrit la lettre.
- FQ2 demande quel est le destinataire de la lettre.
- Il semble donc que les élèves ont facilement compris comment s'organise une lettre, mais aussi qui en est le producteur et quel en est le destinataire.
- FQ5 demande aux élèves de chercher le mot introducteur de l'argument de la maman de Paul (parmi 4 propositions, les distracteurs ayant été par la suite considérés comme trop faciles, donc incitant peu à les choisir).
- FQ6 demande aux élèves de souligner l'opinion de Paul.
- FQ7a demande aux élèves pourquoi Paul a écrit sa lettre et propose quatre possibilités de réponse. Cette question nous semble avoir une parenté avec la question F01 de l'épreuve adaptative, dans la mesure où elle demande de rendre compte d'une bonne compréhension du sens général du texte (ici de la lettre). L'examen des corrélations établies sur le plan statistique montre qu'entre les questions F01 et FQ7a la corrélation est à la fois négative et non significative. Nous pouvons penser que la question *pourquoi* renvoie à une compréhension certes générale du texte, mais qui va plus loin dans la mesure où elle implique une inférence liée à l'intention de l'auteur (ici Paul). La parenté entre F01 et FQ7a n'est donc pas corroborée.

Les trois autres questions restantes sont, quant à elles, particulièrement difficiles :

- FQ3 demande aux élèves de souligner l'opinion de la maman de Paul. Il n'y a donc pas de propositions de réponses à choix, ce qui constitue une modalité de réponse plus complexe. De plus, c'est toute une phrase qui est à souligner et pas seulement un mot. Nous avons toutefois tenu compte des soulignements partiels, dont certains ont été admis comme exacts (et un seul comme partiel : *Elle ne veut pas*).

- FQ4 demande aux élèves de souligner l'argument donné par la maman de Paul. Il faut tout d'abord relever que 10% des élèves n'ont pas répondu à la question, ne l'ayant vraisemblablement pas comprise (le mot « argument » peut avoir fait obstacle à la compréhension). Ensuite, il faut bien reconnaître que l'argument donné n'est pas très facile à identifier, de par son caractère général (*c'est mauvais*). L'indice principal de reconnaissance est grammatical, par la présence de « parce que » comme introducteur de l'argument. Encore faut-il savoir cela...
- FQ7b demande aux élèves de justifier la réponse qu'ils ont donnée à la question précédente en soulignant dans le texte le passage qui explique cette dernière. La tâche ici est complexe, en particulier pour de si jeunes élèves.

Des commentaires à propos des questions difficiles en compréhension

- En ce qui concerne la FQ3, relevons que lorsque les élèves doivent souligner l'opinion de Paul, ils éprouvent moins de difficultés que pour le soulignement de l'opinion de sa maman ; est-ce parce que celle de Paul est exprimée en « je », plus aisé à identifier que « elle » pour sa maman ? Ou parce que la modalisation « je pense », aide davantage à comprendre qu'il s'agit d'une opinion ? Il semblerait en définitive que les indices de reconnaissance de l'opinion ont joué un rôle dans la réussite aux questions davantage que la modalité de la question ou que le fait de savoir ou non ce qu'est une opinion.
- La FQ7b déborde les exigences de la fin de cycle 1. Il y a cependant un gros tiers des élèves qui parviennent à la bonne réponse.

Nous pouvons affirmer, au regard de ces résultats en compréhension, que les élèves sont plutôt bien entrés dans la tâche complexe et ont réussi majoritairement cette partie de compréhension de la lettre de Paul, alors même que son genre est encore peu travaillé dans le cycle 1.

Premiers résultats - En production

Tableau 5.2.2f: Résultats par question en pourcents

%	FP1 + FP5	FP2	FP3	FP4	FP6	FP7	FP8	FP9	FMODIF
Réussite/réussite partielle	1: 91	1: 59 3: 15	1: 60	2: 50 3: 18 1: 13	1: 39 2: 37	1: 81	1: 87	6: 10	1: 27
Non-réussite	0: 9	0: 11	0: 25	0: 6	3: 24	0: 19	0: 13		0: 65 2: 9
Absence		9: 15	9: 15	9: 14				9: 90	

Pour l'ensemble des critères:

code 1 signifie écriture attendue, correcte (sauf FP4 et FMODIF)

code 0 signifie erreur(s) (sauf FMODIF)

code 9 signifie que le texte ne fournit aucune trace correspondant au contenu du critère

FP2: code 3 signifie opinions complémentaires ajoutées, opinion nuancée

FP4: code 1 signifie présence d'un seul argument cohérent

- code 2 : présence de deux arguments cohérents au moins
- code 3 signifie arguments cohérents ; s'y ajoutent un ou plusieurs contre-arguments (par exemple introduits par mais)
- FP6 : code 2 signifie syntaxe partiellement correcte
- code 3 signifie problème de syntaxe
- FP9 : code 6 : ajout d'un contenu supplémentaire
- FMODIF code 1 : modification apportée bénéfique
- code 0 : pas de modification apportée
- code 2 : au moins une erreur ajoutée par une modification du texte

Dans l'ensemble, et compte tenu des critères de correction adoptés, la tâche de production de texte est généralement bien réussie par les élèves. En effet, cinq critères (sur sept en considérant à part FP9 et FMODIF) sont pleinement atteints :

- FP1 et FP5 (regroupés) portent sur le respect du format de la lettre (dont le modèle est donné d'emblée) : l'en-tête et la signature figurent très majoritairement dans les textes des élèves. Par conséquent, les élèves s'adressent bien à Paul, qu'ils tutoient comme l'un de leurs camarades.
- FP2 atteste que les élèves sont capables de donner leur opinion sur le sujet exposé par Paul. Certains vont même jusqu'à nuancer leur opinion (*Moi mon opinion c'est qu'on peut manger des bonbons mais pas tout le temps... (Camille)*).
- FP4 atteste que les élèves sont capables de donner au moins deux arguments allant dans le sens de leur opinion, certains pouvant même en ajouter d'autres, tout aussi cohérents, voire proposer des contrarguments (*Moi je pense que on peut en manger mais pas trop souvent. Parce que ça peut donner des caries et ça peut faire mal au ventre et par contre c'est bon*³⁴. (Lisa)).

Il semble donc que l'activité de mise en situation a effectivement permis aux élèves de comprendre ce que sont opinions et arguments de telle sorte qu'ils sont parvenus à répondre à la lettre de Paul de façon fondée et cohérente. Encore fallait-il donner un contenu et une forme à ces éléments discutés en classe.

- FP7 atteste que les élèves parviennent, dans la plupart des cas (on accepte qu'une ou deux majuscules manquent dans un texte qui comporte plus de cinq phrases ou propositions), à commencer leurs phrases par une majuscule.
- FP8 atteste que les élèves sont capables de copier correctement les mots de leur réservoir de mots (une erreur est tolérée) ; les erreurs d'accord ne sont pas prises en considération ; on ne compte qu'une faute par mot ; une même erreur n'est comptabilisée qu'une fois).

Deux critères ont révélé néanmoins quelques problèmes d'écriture spécifiques.

- FP3 concerne la modalisation de l'opinion, c'est-à-dire le fait d'introduire son opinion par une expression telle que *je trouve que, je pense que, je suis d'avis que, pour moi, à mon avis, je suis pour...* Il n'en demeure pas moins que 60% des élèves ont modalisé leur opinion ce qui peut être lu comme un bon résultat.
- La syntaxe (FP6) n'est pas encore correcte pour plus de 60% des élèves. Ce résultat n'étonne guère, alors que les élèves sont au début de leur apprentissage de l'écriture.

³⁴ L'orthographe a été corrigée.

Par ailleurs, ils sont peu à ajouter du contenu supplémentaire (à ce qui est demandé) dans leur texte (10%).

Enfin, il est difficile à ces élèves de début de 5^e (3P) de prendre du recul par rapport à leur texte et d'y apporter des corrections (formelles). Seul un tiers se lance dans cette opération de correction ; pour les trois quarts d'entre eux, la correction améliore le texte, ce qui n'est pas le cas pour les autres.

Analyse des correspondances multiples

Comme pour l'épreuve adaptative, l'identification des axes de cette analyse s'avère très difficile pour l'épreuve papier.

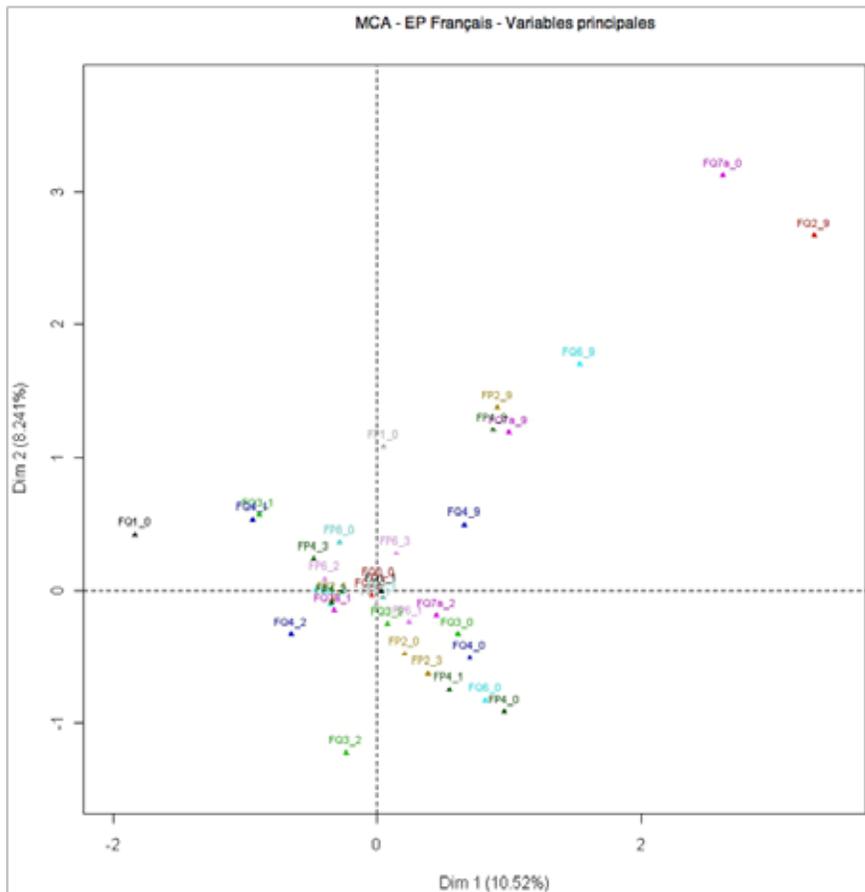


Figure 5.2.2b: Répartition des catégories de variables principales pour l'EP français

Si l'on regarde plus attentivement les zones délimitées par les axes, on peut, grosso modo, constater les points suivants :

- Sur la gauche de l'axe horizontal et dans la partie inférieure de l'axe vertical, les résultats aux questions qui s'y trouvent attestent d'une réussite relativement bonne en compréhension surtout, en production aussi (en particulier pour ce qui touche au format de la lettre, aux opinions et aux arguments), avec des réponses au questionnaire parfois partielles.
- Toujours sur la gauche de l'axe horizontal, mais dans la partie supérieure de l'axe vertical, les résultats montrent une bonne réussite dans ces deux domaines pour les tâches les plus difficiles portant sur les opinions et les arguments. Toutefois la présence des questions et des critères de production suivants nous interrogent : la difficulté d'indiquer quel est l'auteur de la lettre ; une syntaxe, à l'écrit, partiellement maîtrisée et des erreurs de copie de mots issus du réservoir de mots. Il apparaît ainsi que les élèves se situant dans ce secteur ont à la fois dominé les tâches les plus complexes et buté sur des éléments de mise en texte ou sur une question facile liée au format de la lettre (98% de réussite) !
- La partie à droite de l'axe horizontal et supérieure de l'axe vertical regroupe une majorité de non-réponses, accompagnées de résultats faux pour la reconnaissance du destinataire de la lettre de Paul, la prise en considération du format de la lettre en production ; de plus, la syntaxe est problématique.
- Enfin, la partie de droite de l'axe horizontal et inférieure de l'axe vertical présente une majorité d'échecs aux questions de compréhension et de production, à l'exception de ce qui touche à la syntaxe et à la copie de mots, bien réussi.

Discussion des résultats et hypothèses en lien avec la production

- L'analyse des correspondances multiples montre que dans la partie supérieure de l'axe vertical, sur la gauche de l'axe horizontal, donc dans la zone des questions les plus complexes, réussies, se trouvent quelques critères de la production écrite portant sur la mise en texte, tels que la syntaxe et la copie de mots, médiocrement réussis. Nous faisons l'hypothèse que pour de très jeunes élèves occupés à comprendre la situation, à écrire un contenu complexe, la surcharge cognitive peut s'avérer telle que la mise en texte elle-même est laissée pour compte ou source de difficultés. De plus, on observe que dans cette même zone se trouve une variable globale relative à la correction du texte lors de la relecture du texte qui montre que des élèves qui réussissent bien la tâche complexe peuvent néanmoins, au moment de la relecture, ajouter dans leur texte une erreur portant sur la forme.
- Dans ce sens, comme dans l'épreuve adaptative, la phrase écrite n'est pas encore véritablement reconnue et stabilisée, et les mots de la phrase sont, pour certains, mal segmentés. Un travail grammatical reste à être mené, dont une part en lien avec l'écriture même d'un texte, afin que les élèves comprennent comment articuler leurs savoirs sur la langue à la production textuelle. De plus, l'agencement des idées, leur organisation en texte, demandent un important effort cognitif aux élèves avec pour conséquence de reléguer au second plan la mise en texte elle-même (syntaxe, orthographe...).
- L'absence d'ajouts d'un contenu original dans le texte n'est là encore pas étonnante, la tâche de base réclamant déjà beaucoup de la part de si jeunes élèves (maîtrise du contenu et mise en texte). De plus, il est plausible que des élèves ne voient pas l'utilité de faire davantage que ce que demandent les consignes. Il reste

que ces ajouts, observés dans plusieurs textes, nous semblent des signes d'une certaine aisance ou d'une certaine créativité dans la production de texte.

- Enfin, la difficulté de revenir au texte, de le relire et d'y repérer des erreurs et des améliorations à lui apporter n'est guère surprenante non plus car elle demande de l'élève une posture de distanciation par rapport au texte, peu familière pensons-nous à ce moment de la solarité. Cependant, cette capacité de correction est déjà présente chez certains ; selon nous, elle pourrait être développée déjà au cycle 1.

... et en lien avec l'EP dans son ensemble

Quelques questions, provenant aussi bien de la partie compréhension que de la partie production, différencient très clairement les élèves qui réussissent très bien la tâche complexe de ceux qui éprouvent des difficultés dans les deux domaines ; ce sont les questions qui touchent à la reconnaissance et à la formulation des opinions et des arguments. Les très bons « compreneurs » et « producteurs » peuvent désigner l'opinion de Paul et de sa mère, puis, dans leur texte, fournir la leur et donner au moins deux, voire trois arguments à l'appui (FQ3, FQ4, FP2, FP4). S'ajoute à cela la capacité, chez ces bons scripteurs, de modaliser leur opinion (je pense que... , etc.). En compréhension, deux autres questions, complémentaires, jouent également un rôle d'indicateurs de la bonne réussite à la tâche complexe : savoir indiquer correctement la raison pour laquelle Paul écrit sa lettre et, ce qui est plus difficile, indiquer par un soulignement le passage de la lettre qui corrobore cette raison évoquée (FQ7a et FQ7b). Nous faisons l'hypothèse (car nous n'avons pas pu vérifier cette assertion) que les élèves qui répondent avec succès à ces questions devraient réussir l'ensemble de la tâche complexe, c'est-à-dire devraient être capables de comprendre les enjeux de la situation, les contenus qu'elle met en scène, et de traduire cette double compréhension dans un texte écrit, même si les outils formels ne sont pas encore tout à fait maîtrisés.

Ajoutons encore deux hypothèses liées à des questions ou critères spécifiques : nous nous demandons, à la suite des constats précédents, si une des caractéristiques déterminantes de la réussite à la tâche complexe ne reposerait pas sur la capacité des élèves au-delà de la simple compréhension du texte, à s'en distancer suffisamment pour pointer des éléments explicatifs de certains passages, ce qui ressort par exemple des questions FQ7a et b. En production, cette capacité de distanciation se manifesterait dans le critère FP3 sur la présence d'une modalisation pour asseoir son opinion. On pourrait y ajouter la capacité d'ajouter un contrargument plausible au texte argumentatif (FP4, modalité 3) qui attesterait déjà d'une certaine maîtrise de l'échange argumentatif sur un sujet simple, adapté aux élèves évidemment. Nous trouverions donc intéressant de vérifier si des corrélations existent entre ces questions et critères (car nous n'avons pas réalisé ces corrélations). Dans le même sens, il nous apparaît que FQ7a demande des élèves une compréhension des intentions de Paul, donc de comprendre le but général de la lettre qu'il a écrite ; dans ce sens, nous faisons l'hypothèse d'une parenté entre cette question et la question F01 de l'épreuve adaptative demandant une compréhension globale du conte. Là encore, une mise en corrélation de ces questions (encore à faire) pourrait confirmer notre hypothèse.

Clusters d'élèves et types de difficultés

La proposition de *clusters* issue de l'analyse des correspondances multiples ne permet guère de constats saillants. Nous avons donc préféré partir de nos observations qualitatives pour poser – sous forme d'hypothèses – les éléments suivants.

Nous avons constaté que la partie production a été généralement – et en fonction de nos critères d'évaluation – mieux réussie que la partie compréhension. On compte en effet pour 33 élèves qui éprouvent des difficultés dans la partie compréhension de la lettre de Paul, 8 élèves seulement qui, en plus, peinent à écrire la lettre de réponse. Seuls 6 réussissent la partie compréhension mais échouent la partie production; nous y reviendrons. Il faut redire qu'après l'évaluation de la compréhension de la lettre de Paul par questionnaire, la mise en situation s'est poursuivie par la correction du questionnaire; ainsi, lorsque les élèves ont ensuite écrit leur lettre-réponse, ils avaient en principe les clés de compréhension du contenu de la lettre de Paul.

Nous l'avons dit, 6 élèves, dont 5 appartenant à une même classe – qui par ailleurs a très bien réussi la tâche de production – réussissent la partie compréhension mais connaissent des difficultés dans la partie production (au moins 4 critères sur les 7 posent problème). Plus précisément, 5 de ces élèves sur 6 éprouvent des difficultés à modaliser leur opinion, en l'introduisant par une formule telle que *à mon avis*; 4 sur 6 oublient la plupart des majuscules au début de leurs phrases et leur texte présente une syntaxe peu satisfaisante; de plus, ils ne parviennent pas à fournir au moins deux arguments à l'appui de leur opinion. La moitié d'entre eux font des erreurs de copie aux mots tirés de leur réservoir de mots et ne donnent pas leur opinion. 2 élèves se trompent à propos du format de la lettre (en-tête et/ou signature). Finalement, ce sont sur des aspects formels que les erreurs sont le plus nombreuses.

Cela nous conduit tout naturellement à envisager maintenant les liens que l'étude statistique a mis en évidence entre les deux épreuves, l'épreuve adaptative et l'épreuve papier sur la tâche complexe.

En fonction de ce qui précède, nous nous demandons en effet, dans quelle mesure les élèves qui ont éprouvé des difficultés de compréhension dans l'EA en ont également connu dans l'EP; nous nous demandons encore si les élèves qui ont plutôt de la peine à répondre aux questions relatives au fonctionnement de la langue dans l'EA se trouvent en difficulté dans les questions techniques de compréhension, mais surtout dans la mise en texte de leur réponse à Paul. Enfin, nous voudrions savoir comment se comportent dans l'EA les quelques élèves qui ont plus particulièrement échoué à la production de texte.

ÉVALUATION EN DEUX PHASES (EA+EP)

En nous appuyant sur les hypothèses et les questions que nous venons de formuler, nous allons dans un premier temps examiner les résultats statistiques que nous avons obtenus.

Analyse des correspondances multiples

Nous allons procéder comme précédemment, en tentant d'identifier les zones délimitées par les axes de l'analyse des correspondances multiples, mais aussi en recherchant le sens que peuvent prendre certaines variables se caractérisant par une certaine proximité dans l'une ou l'autre de ces zones.

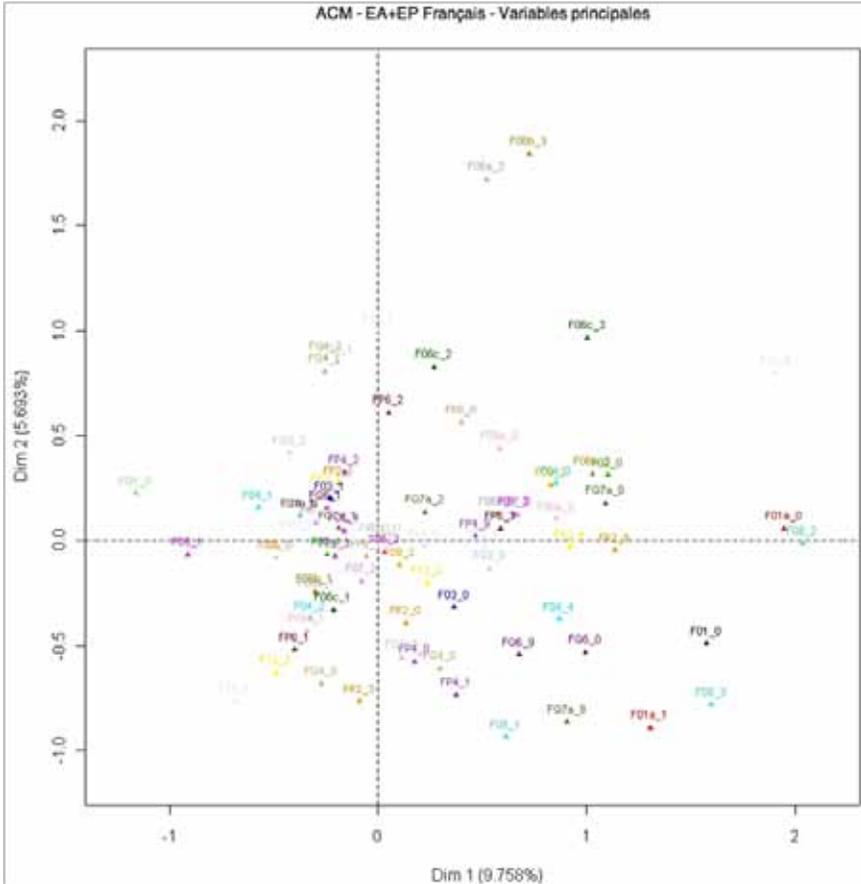


Figure 5.2.2c: Répartition des catégories de variables principales pour l'EA et l'EP français

Si, comme dans les analyses précédentes, les bons résultats se démarquent des moins bons, sur l'un des axes on peut, pour ces premiers, délimiter deux ensembles de variables :

- un ensemble regroupant des catégories de la réussite à des tâches majoritairement en lien avec le fonctionnement de la langue ou la mise en texte en production. Ainsi, dans l'EA, la question sur la reconnaissance d'anaphores placées dans un court texte ; les questions touchant à la segmentation de phrases en mots ; celle relative à la distinction phonèmes-graphèmes ; et celle touchant au

découpage d'un court texte en phrases. S'y ajoutent, entre autres, la qualité de la syntaxe et la copie correcte de mots dans la production écrite;

- un ensemble regroupant des catégories de la réussite à des tâches centrées prioritairement sur la compréhension de texte, que ce soit dans l'EA ou dans l'EP. Dans la première de ces épreuves, relevons la question de compréhension globale du texte; celle sur la reconnaissance du genre de texte; celle sur l'identification de locuteurs à partir d'extraits de dialogue; celle demandant d'indiquer pour plusieurs propositions du texte si elles étaient vraies ou fausses; et enfin la question sur la mise en relation de plusieurs pronoms personnels à des formes verbales. Du côté de l'EP, à propos de la partie compréhension de la lettre de Paul, se trouvent dans cette zone les questions sur la destination de la lettre; sur l'opinion de la maman de Paul et son soulignement; sur le soulignement de l'argument donné par la maman de Paul; sur le soulignement de l'opinion de Paul; sur la reconnaissance de la raison de l'envoi de la lettre par Paul. En production écrite, on observe encore la présence d'une ou plusieurs opinions cohérentes dans le texte écrit.

En ce qui concerne les zones rassemblant des catégories de non-réussite aux tâches, on constate le pendant des groupes précédents, mais en négatif. Il y aurait donc grosso modo deux grandes catégories de difficultés repérées: celles portant essentiellement sur le fonctionnement de la langue et la mise en texte en production d'une part, celles portant sur la compréhension, que ce soit dans l'EA ou dans l'EP d'autre part.

Clusters d'élèves et types de difficultés

Faisons un pas de plus et examinons les *clusters* proposés à partir de l'analyse des correspondances multiples. Trois *clusters* ont été déterminés. Le premier d'entre eux (*cluster A*) regroupe 65 élèves dont l'ensemble des résultats (dans l'EA et dans l'EP) peuvent être considérés comme bons (sachant, ainsi que nous l'avons vu dans les résultats précédents, que certaines questions ont été difficiles, dépassant pour quelques-unes d'entre elles les exigences de la fin du cycle 1). Comme nous l'avons déjà souligné précédemment, et au vu de la distribution des variables principales dans ce *cluster*, ce qui caractérise particulièrement les meilleurs élèves est la capacité d'utiliser la langue à des fins d'expression (modaliser son opinion, reconnaître le rôle de « parce que », au début d'un argument, trouver dans le texte ce qui rend compte d'une opinion ou d'un argument, savoir justifier une réponse à l'aide du texte, etc.).

Un deuxième *cluster* (*cluster B*) met en évidence des difficultés plus particulièrement en lien avec le fonctionnement de la langue (segmentation de phrases en mots, reconnaissance des anaphores, la distinction phonèmes-graphèmes, une syntaxe lacunaire en production écrite), alors que bien des questions de compréhension ou des critères de production de texte sont réussis. 15 élèves appartiennent à ce *cluster*.

Quant au troisième *cluster* (*cluster C*), il laisse voir, chez les 19 élèves qui le composent, des difficultés majeures en compréhension de texte (compréhension globale, compréhension de qui dit quoi, repérage difficile des opinions et des arguments), mais aussi quelques difficultés dans le fonctionnement de la langue, en particulier la reconnaissance des phonèmes-graphèmes, l'identification de phrases dans un texte, la mise en relation de pronoms personnels avec des formes verbales, une syntaxe peu maîtrisée en production de texte.

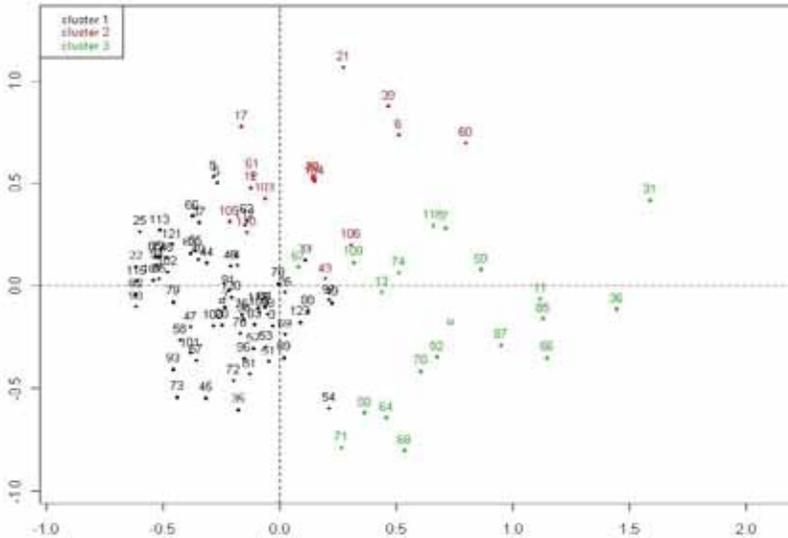


Figure 5.2.2d: Répartition des catégories de variables principales pour les EA et EP français

Analyses complémentaires

Nous avons complété cette analyse en *clusters* sur la base de deux nouvelles hypothèses. Le but de ces nouvelles investigations, réalisées de manière empirique et qualitative, était l'affinement des catégories de difficultés précédemment identifiées.

Tout d'abord, nous sommes partis de l'hypothèse que la question 1 de l'EA (F01) (compréhension globale du texte) peut entraîner un ensemble d'erreurs spécifiques. Nous avons dénombré 18 élèves qui échouent à cette question et avons examiné leur parcours dans les deux épreuves. Il s'avère que ce groupe (que nous appellerons groupe 1) correspond largement au troisième *cluster* (*cluster C*) mis en évidence ci-dessus. Il est composé d'une majorité de garçons, ayant tendance à lire le texte de départ trop vite (N=11 élèves) (une minute ou moins) et quelques-uns (N=5) très lentement (plus de 5 minutes). Beaucoup vont relire le texte de *Tibili le petit garçon qui ne voulait pas aller à l'école* à l'occasion d'une question ou d'une autre, mais leur relecture est peu productive et conduit le plus souvent à un échec à la question. La moitié environ de ces élèves ne réussissent que 0 à 4 items de l'épreuve adaptative. Il ressortirait ainsi que l'échec à la question F01 constitue un indicateur d'une compréhension difficile du texte.

La deuxième hypothèse met en avant qu'une importante difficulté à écrire un texte pourrait signifier d'importantes lacunes dans plusieurs domaines du français. Un nouveau groupe (groupe 2), rassemble des élèves – peu nombreux (N=9) – qui présentent des difficultés en production écrite, tâche complexe par ailleurs généralement bien

réussie par l'ensemble des élèves. Nous voulions connaître le parcours de ces élèves dans les deux épreuves. Or, nous constatons (pour en tout cas 7 d'entre eux) qu'ils éprouvent des difficultés de base dans la segmentation des phrases en mots, dans la reconnaissance de phonèmes-graphèmes, accompagnées de difficultés de compréhension : trouver la raison d'une action relatée dans le texte, associer des dialogues à des personnages et des anaphores aux sujets correspondants. S'ajoute la difficulté – courante encore en fin de cycle 1 – à reconnaître des phrases dans un texte écrit en continu.

S'ajoute à ces difficultés une difficulté liée à la tâche complexe, celle de comprendre ce que signifie une opinion et la manière dont elle s'exprime dans un texte. Il en est de même pour l'argumentation. La justification d'une réponse donnée est également particulièrement difficile pour ces élèves, d'autant plus que cette tâche semble dépasser les possibilités de fin de premier cycle (un peu plus du tiers de réussite seulement pour l'ensemble de la population). Il est intéressant de remarquer que la difficulté rencontrée en compréhension à propos de l'opinion, mais aussi des arguments, se répercute, pour ces élèves, sur la production. De manière globale, ces élèves connaissent donc des difficultés suffisamment importantes pour faire obstacle à l'écriture d'un texte qui pourraient encore toucher au déchiffrement, mais certainement à la compréhension de texte.

Enfin, nous nous sommes demandé s'il était tout de même possible de réussir la tâche complexe (EP) tout en éprouvant des difficultés dans l'EA, ce qui *a priori* peut paraître surprenant, les compétences requises dans l'EP étant d'un niveau plus élevé que la majorité de celles de l'EA (de tels constats ont toutefois été relevés dans l'ouvrage de Roegiers de 2004). Nous avons alors repéré 9 élèves (groupe 3) qui réussissent avec succès la tâche complexe de l'EP, aussi bien la partie compréhension que la partie production et qui cependant, lorsqu'on regarde leurs résultats dans l'EA, rencontrent quelques problèmes : ceux-ci sont relatifs à la mise en lien d'éléments dialogués avec les personnages qui parlent, à la reconnaissance d'anaphores dans un court texte, à la segmentation de phrases en mots (deux phrases sur trois), à l'identification de phrases dans un court texte. Pour la moitié de ces élèves, la reconnaissance phonèmes-graphèmes pose également problème. Finalement, ces 9 élèves connaissent passablement de difficultés dans l'EA. En cela, ils rejoignent leurs camarades du groupe précédent, sauf que, eux, semblent comprendre le texte de la lettre de Paul. Ils sont donc entrés dans la situation présentée et nous pouvons penser qu'ils ont largement su bénéficier de l'activité préalable. Il reste qu'il est possible, malgré des lacunes dans le domaine du fonctionnement de la langue, de réussir une tâche complexe de production de texte. Nous nous demandons alors si la production ne serait pas davantage corrélée à la compréhension de texte et à celle de la situation propre à la tâche demandée plutôt qu'aux connaissances sur le fonctionnement de la langue.

Catégorisation des questions et critères des épreuves EA et EP

À propos de la tâche complexe, que ce soit pour la compréhension ou pour la production de texte, nous avons vu, au travers des tables de spécification, que les critères retenus se regroupaient autour de catégories plus larges, traversant ces domaines, que nous avons appelées composantes. Il nous a semblé en effet intéressant de procéder à un regroupement des critères de l'EP pour tenter à la fois de mieux cibler des niveaux de compétences plus globales liées aux tâches que nous y avons définies et

de voir si, au travers de cette catégorisation, nous observerions une stabilité des résultats selon les composantes définies.

Dans le même sens, Roegiers (2004) se demande dans quelle mesure les critères resteraient semblables dans une discipline donnée.

« Les critères sont relatifs à une compétence dans une discipline donnée. Ils sont donc les mêmes, quelle que soit la situation que l'on apprécie en relation avec cette compétence. Mais il y a encore plus : en général, on recourt toujours aux mêmes types de critères dans une discipline donnée, voire même dans un champ disciplinaire donné. Les critères sont donc non seulement stables par rapport à une compétence donnée, mais également par rapport à une discipline donnée » (p. 199)

Roegiers reconnaît qu'il peut y avoir des variations par rapport à cette stabilité des critères. Ainsi, en ce qui nous concerne, nous pensons qu'une composante supplémentaire devrait être ajoutée pour les degrés suivants de la scolarité, celle de la planification (cf. au point 4.2.3, l'origine et le descriptif des composantes retenues).

Dans cette perspective, nous avons réalisé une catégorisation des critères retenus pour la correction de l'épreuve EP en fonction des trois composantes établies pour l'évaluation de la tâche complexe et présentes dans les tables de spécification.

À l'instar de Roegiers, nous nous sommes encore demandé dans quelle mesure ces composantes pouvaient avoir un sens pour tout ou partie des questions de l'EA, plus particulièrement les questions de compréhension de texte. Il nous semblait en effet que cela pouvait nous donner des clés supplémentaires pour l'interprétation de nos données. Il reste que, pour un certain nombre de questions portant sur le fonctionnement de la langue ne renvoyant pas à un texte, même très court, il apparaît non pertinent de les placer dans la rubrique intitulée textualisation. C'est pourquoi, nous avons retiré les questions F06, F09 et F12 de cette catégorisation.

Dans ces trois catégories, nous avons placé questions et critères en précisant de quelle épreuve ou partie d'épreuve ils émanaient.

Appréhension du contexte :

Dans l'EA – F03³⁵

Tâche complexe de compréhension de texte – FQ1 – FQ2 – FQ7a – FQ7b

Tâche complexe de production de texte – FP1

Traitement du contenu :

Dans l'EA – F01 – F02 – F04 – F07

Tâche complexe de compréhension de texte – FQ3 – FQ4 – FQ6

Tâche complexe de production de texte – FP2 – FP4

Textualisation :

Dans l'EA – F05 – F10

³⁵ Les questions soulignées concernent des tâches dépassant les exigences requises pour des élèves de fin de premier cycle.

Tâche complexe de compréhension de texte – FQ5

Tâche complexe de production de texte – FP3 – FP6 – FP7 – FP8

Nous faisons ainsi l'hypothèse que cette catégorisation des items va offrir une base supplémentaire d'interprétation des résultats obtenus, générer des liens entre les tâches et nous aider à affiner les *clusters* d'élèves et les types de difficultés définis précédemment.

Si nous reprenons les *clusters* et les groupes décrits plus haut à propos des résultats dans les deux épreuves EA et EP, nous pouvons compléter leurs caractéristiques de la manière suivante :

Les élèves du *cluster* B rencontrent leurs principales difficultés dans la composante de la textualisation. Comme pour les élèves du groupe précédent (groupe 3), ils sont capables d'appréhender le contexte d'une tâche complexe.

Ceux du *cluster* C (et du groupe 1) éprouvent principalement des difficultés de compréhension qui attesteraient de lacunes dans le traitement du contenu du texte. Cela ne signifie pas pour autant qu'ils peinent à appréhender le contexte d'une tâche complexe.

Les élèves du groupe 2 montrent des lacunes dans les trois composantes ; en particulier, ils ne semblent pas percevoir les enjeux de la tâche complexe, peinant à comprendre ce que sont opinions et arguments et comment les associer pour répondre aux exigences de la tâche. Tout s'imbrique en fait ; c'est à la fois la difficulté de saisir les paramètres du contexte et celle de comprendre les contenus que ce dernier implique et engendre qui se combinent, accompagnées de difficultés liées à la textualisation.

Il en est tout autrement des quelques élèves qui réussissent bien l'EP mais moins bien l'EA (groupe 3). Dans ce cas, les élèves perçoivent tout à fait bien le contexte, ses enjeux, et répondent aux exigences des tâches de compréhension et de production. Et pourtant, en particulier dans la partie de l'épreuve adaptative portant sur des connaissances grapho-phonétiques, mais aussi sur des points relatifs aux reprises anaphoriques ou à la segmentation d'un court texte en phrases, ils éprouvent encore des difficultés.

Ces quelques constats pourraient avoir une influence sur les remédiations envisagées par les enseignants. On pourrait en effet imaginer des activités spécifiques en fonction des composantes identifiées. Mais ce travail est encore à réaliser, et nous ne nous avancerons pas davantage sur ce sujet ici.

En conclusion

Nous voulons tout d'abord souligner combien notre investigation – par le biais de l'EA comme de l'EP – est large, cherchant à toucher de nombreuses attentes fondamentales relatives à deux objectifs d'apprentissage de grande amplitude. Nous la considérons comme une exploration attestant de la difficulté qui existe encore de comprendre exactement comment s'articulent les différents niveaux de compétence lorsqu'ils touchent à des objets aussi complexes que la compréhension ou la production de textes (qu'elles soient orales ou écrites, d'ailleurs). Quant à savoir en sus comment les éléments relatifs au fonctionnement de la langue s'imbriquent à titre de

ressources dans ces domaines, bien des investigations restent à faire pour y parvenir. Nous avons néanmoins tenté d'y voir plus clair dans ces relations.

Tout d'abord, il nous paraît tout à fait important de couvrir, par l'évaluation, ce qui touche aux connaissances et procédures liées aussi bien au fonctionnement de la langue qu'aux stratégies de lecture (même si de ce point de vue notre travail est lacunaire) et ce qui touche aux tâches complexes de compréhension-production. En effet, dans une logique de l'évaluation de compétences, ces différents domaines sont liés, puisque nombre de connaissances et procédures sont des ressources pour la résolution de la tâche complexe.

Les résultats montrant que ce qui est considéré comme complexe n'est pas forcément moins bien réussi que ce qui relève de compétences de niveaux moins complexes, et même, qu'il est possible de mieux réussir les tâches complexes que les questions plus ciblées et portant sur des connaissances ou procédures considérées *a priori* comme plus simples, nous estimons important :

- d'aborder en classe – éléments figurant dans le PER au même titre que les apprentissages de connaissances – des tâches complexes dès le plus jeune âge.
- De faire comprendre et produire des textes de tous genres (la production d'un texte argumentatif de l'EP, certes préparée par le biais d'une activité préalable, a été plutôt bien réussie, constituant un bon pronostic pour la suite des apprentissages) aux élèves du premier cycle, évidemment adaptés à leur âge (choix des textes à lire, choix des contenus, des situations de production, etc.).

Nous avons vu encore l'importance de la compréhension globale d'un texte comme indicateur d'une bonne compétence en compréhension. Cette piste, tant du point de vue didactique qu'évaluatif, reste à suivre. À ce propos, et à l'appui de cette affirmation, les résultats de l'étude PISA 2009 en littératie pour la Suisse romande (Nidegger, 2011) montrent, entre autres, combien certaines stratégies de lecture, parmi lesquelles la capacité de synthèse, jouent un rôle positif dans de bonnes compétences en compréhension de la lecture :

« Dans l'ensemble, ces stratégies confirment l'importance de saisir les visées de la lecture, de tenir compte des besoins du lecteur par rapport au texte à lire, et la nécessité de disposer de moyens de vérification de la compréhension, en particulier par la reconnaissance des éléments les plus importants du texte lu et la capacité de les résumer » (p. 137).

Certes, ces propos touchent des jeunes en fin de scolarité, mais nous avons vu que le PER mentionne déjà, dans les attentes fondamentales et pour la fin du cycle 1, la compréhension globale d'un texte court et, notamment d'un texte narratif. Cette compréhension passe par celle des intentions de l'auteur, ainsi que par l'identification des idées principales d'un texte en fonction des buts assignés à la lecture de celui-ci ; (signalons encore ici une étude de Dubois et Wirthner, 1990, sur l'objectif *saisir l'idée principale d'un texte* ; elle met en évidence la manière dont des élèves de 5^e appréhendent la signification de l'idée principale d'un texte et de ses liens avec le résumé).

Nous voulons enfin terminer par quelques remarques relatives aux remédiations envisageables à la suite de nos investigations.

Pour l'apprentissage de la production de textes de genres divers, il existe en Suisse romande un moyen à la disposition des enseignants de tous les degrés de la scolarité obligatoire : *S'exprimer en français*. Ce moyen propose des séquences didactiques pour tous les regroupements de genres. Certes, il aurait besoin d'être actualisé (il a plus de dix ans d'âge) et rendu plus souple d'utilisation, mais il est en adéquation avec les exigences du PER. Il est ainsi à même de préparer les élèves à résoudre des tâches de production telles qu'envisagées dans notre travail.

En ce qui concerne la compréhension écrite, il nous semble qu'un travail sur les stratégies de lecture devrait être proposé pour faciliter la compréhension de textes. L'idéal serait que ces stratégies soient adaptées aux différents genres de textes. Pour l'instant, le seul matériel existant en Suisse romande est *Lector & Lectrix* (Cèbe et Goigoux, 2009), prévu pour des élèves du cycle 2 et du début du secondaire et pour des textes narratifs uniquement. Il existe par ailleurs, et cela s'est particulièrement développé depuis les premiers tests PISA de 2000, de nombreux dispositifs d'apprentissage développant la compréhension, l'interprétation de la lecture ; non officiels, il dépend alors des enseignants de les introduire dans leur classe...

Quant au travail sur le fonctionnement de la langue, il peut s'appuyer sur un important matériel proposé en Suisse romande. Provenant de France, ce matériel a été adapté à la situation locale ; mais il n'en demeure pas moins que des incohérences (terminologiques en particulier) peuvent subsister selon les manuels et les éditions.

Il est certain que nos épreuves sont perfectibles, et des propositions d'amélioration ont déjà été faites dans le cadre de nos groupes de travail. Certaines d'entre elles figurent d'ailleurs dans l'annexe 8.3. C'est pourquoi nous insistons sur le caractère exploratoire de notre investigation dans les classes, dont le but principal est alors de fournir un matériel à discuter, à modifier si nécessaire et à développer. De plus, il existe une grande expérience des évaluations externes dans les cantons. La suite des travaux devrait reposer également sur cette expérience tout en prenant en considération à la fois le PER et le matériel que nous venons de constituer.

5.2.3 Les résultats de l'épreuve de mathématiques – cycle 1

L'ÉPREUVE ADAPTATIVE (EA)

L'épreuve informatisée s'est concentrée essentiellement sur la mesure de ressources à mobiliser dans les tâches complexes devant permettre de fournir une information sur l'objectif d'apprentissage *Poser et résoudre des problèmes pour construire et structurer des représentations des nombres naturels* et les attentes fondamentales pour la fin du cycle 1 (cf. 4.3). Cette épreuve est constituée de 11 tâches principales et de 5 tâches complémentaires, soumises aux élèves selon leur réussite à certaines tâches principales.

Contrairement à l'EA en français, nous n'avons pas défini si certaines catégories de réponses fausses pouvaient attester toutefois d'une réussite – même partielle – à l'item, mais nous avons construit les catégories d'erreurs en tentant de repérer certaines difficultés que nous pourrions interpréter et donc auxquelles nous pourrions remédier. Afin toutefois de permettre une vue rapide sur les résultats des élèves, nous avons réparti, dans le tableau 5.2.3, les pourcentages d'élèves selon chaque catégorie

de réponses. Les numéros de tâches comportant un « a » renvoient à des tâches complémentaires soumises seulement aux élèves lorsque la tâche précédente est ratée (en gris clair M01a, M02a, M04a, M05a), ou exceptionnellement lorsqu'elle a été réussie (en gris foncé – M06a) (voir les tâches dans l'annexe 8.4).

Tableau 5.2.3 : Résultats par question en pourcents

%	M01	M01a	M02	M02a	M03	M04	M04a	M05
Code 1*	77%	42%	60%	35%	29%	59%	38%	48%
Code 2					11%	20%	9%	13%
Code 3					17%	10,5%	7%	7%
Code 4					43%	10,5%	46%	32%
Code 0**	23%	58%	40%	65%				

* le code 1 représente la bonne réponse

** le code 0 représente, lorsque les réponses sont dichotomiques, une mauvaise réponse

Pour M03 : Code 2 = réponses justes mais il en manque;

Code 3 = certaines réponses fausses parmi 6, 36 ou 130;

Code 4 = certaines réponses fausses parmi 41, 48, 102, 197 ou 200

Pour M04 : Code 2 = réponse dans $\{45; \dots; 49\} \setminus \{47\}$;

Code 3 = réponse dans $\{20; \dots; 80\} \setminus \{45; \dots; 49\}$;

Code 4 = autre valeur

Pour M04a : Code 2 = réponse 37;

Code 3 = réponse dans $\{42; \dots; 52\} \setminus \{47\}$;

Code 4 = autre valeur

Pour M05 : Code 2 = réponse 4;

Code 3 = réponse 30;

Code 4 = autre valeur

%	M05a	M06	M06a	M07	M08	M09	M10	M11
Code 1*	68%	44%	96%	64%	38%	96%	93%	69%
Code 2	15%	42%		21%	8%			11%
Code 3	17%	14%		15%	4%			7%
Code 4					9%			8%
Code 5					10%			5%
Code 6					15%			
Code 7					16%			
Code 0**			4%			4%	7%	

Pour M05a : Code 2 = réponse 3 ou 20;

Code 3 = autre valeur

Pour M06 Code 2 = dix billes et/ou dix boîtes déposées;

Code 3 = autre erreur;

le code 1 comprend également la réponse avec 7 boîtes et 9 billes (et pas 4 boîtes et 2 billes), le distracteur ayant été considéré trop difficile

Pour M07 : Code 2 = réponses dans $\{x = 10n + 7 \mid n \in \mathbb{N}\}$ (y compris 47 ou 77);

Code 3 = une réponse au moins ne finit pas par 7

Pour M08 : Code 2 = réponse 5 ou 7;

Code 3 = réponse 60 ou 70;

Code 4 = réponse 10;

Code 5 = réponse 68;

Code 6 = réponse 100;

Code 7 = autre valeur

Pour M11 :

Code 2 = mauvais choix de sacs mais ordonnés;

Code 3 = bon choix de sacs mais 81 et 84 sont inversés;

Code 4 = mauvais choix de sacs et ordre inversé (le plus petit sur marche 1);

Code 5 = autre réponse;

le code 1 comprend également les cas où les sacs sont les bons mais une erreur de captage a été repérée

Les **tâches M01 et M01a** ont été, après coup, considérées en décalage par rapport à la mesure recherchée; en effet, elles demandent des calculs (ou le recours au répertoire mémorisé) qui ne sont pas directement en lien avec la compréhension du système décimal, objet qui nous a paru central dans cette épreuve. Il n'en demeure pas moins que nous avons été surpris du nombre d'élèves éprouvant des difficultés à composer 10 à partir de nombres donnés (reconnaissance des complémentaires à 10); toutefois, on peut considérer ici que la consigne laisse une marge d'interprétation et il pourrait être utile de constituer d'autres catégories si la tâche devait être conservée, ou, plus simplement, de préciser sa consigne. La tâche M01a mériterait également d'être améliorée: en effet, le nombre à constituer est passé de 10 dans la tâche M01 à 13 dans celle-ci, impliquant de l'élève d'autres procédures et combinaisons des nombres rendant cette tâche certainement plus difficile que la M01.

La **tâche M02a** annonçait aux élèves qu'ils avaient mal choisi les sacs dans la question M02. Le taux de réussite restant bas, on peut supposer que les élèves ne se souviennent pas forcément des nombres choisis auparavant et peuvent recréer leur erreur, voire qu'une certaine lassitude peut les mener à vouloir passer rapidement à la question suivante. C'est également le cas de la question M04a où le dénombrement des billes peut prendre beaucoup de temps; or toute erreur exigeait de l'élève qu'il recommence, ce qui pourrait être repensé si des tâches de remédiation étaient constituées. En effet, les tâches complémentaires exigent des élèves de refaire parfois une tâche sur laquelle ils ont pu connaître des difficultés et, malgré une aide apportée, cela revient à exiger d'eux une persévérance dont ils ne disposent pas forcément.

La **tâche M05a** est un exemple d'adaptation de la tâche M05, où les boîtes sont représentées ouvertes et où la difficulté liée au contenant semble avoir disparu. Elle permet notamment de constater la résistance des représentations que peuvent se créer certains élèves. Il est intéressant de constater que parmi 52% d'élèves qui échouent à la tâche M05, 32% encore se trompent sur la tâche M05a (soit 22 élèves sur 69) alors que toutes les billes sont désormais visibles et qu'un simple comptage pourrait suffire.

Il est également intéressant de constater que tous les élèves qui réussissent la tâche M05 ne réussissent pas la M06 qui présente une situation similaire (soit 11 élèves sur les 68 qui ont réussi la M05). 15 élèves ayant dû faire la M05a réussissent la M06, dont 12 ont réussi la M05a et avaient un code 4 à la question M05 (signifiant une réponse hors propos *a priori*); ces élèves-là montrent visiblement une capacité à adapter leur représentation, initialement erronée, à partir des indices dont ils disposent.

La **tâche M06a** n'a été présentée qu'aux élèves ayant réussi la tâche M06 et le taux de réussite de 96% est sans appel: la bonne compréhension d'une décomposition d'un nombre en dizaines et unités permet ensuite une extension à une décomposition d'un nombre comportant des centaines. Il aurait été intéressant de soumettre cette tâche directement à tous les élèves puis, dans le cas d'un échec, de présenter la tâche M05,

permettant ainsi d'analyser la simplification que le passage d'une manipulation à des dizaines peut produire.

Le croisement des réponses entre la **M03** et la **M07** est intéressant car, alors que nous pensions faire appel à une compétence relativement similaire (mais pas une compétence complexe à proprement parler), nous constatons que :

- Le taux de réponses correctes est bien supérieur pour la tâche M07 que pour M03 (64% contre 29%) : ceci tend à confirmer une hypothèse faite lors de la construction de l'épreuve, c'est-à-dire que la tâche M07 est effectivement plus simple puisque le nombre de réponses est limité, connu et que les nombres placés dans l'ordre sont une aide importante au repérage.
- Seuls 27 élèves parmi les 85 qui réussissent la tâche M07 ont également réussi la tâche M03, laissant 12 élèves (sur les 39 élèves qui ont réussi la M03) ne parvenant pas à en réaliser une plus simple, alors qu'ils ont réussi une tâche plus compliquée relevant d'une compétence similaire.

Cette simple comparaison nous paraît très importante en terme de contraintes : d'une part, une épreuve devrait présenter plusieurs tâches aux élèves permettant d'extraire des critères comparables (Roegiers, 2004, p. 188) ; d'autre part, le droit de se tromper est à prendre en compte dans l'ensemble de l'épreuve, d'où cette règle des 2/3 qui est suggérée dans l'ouvrage de Roegiers (2004, p. 174).

La **tâche M08** reprend l'idée de boîte contenant 100 billes, à l'instar de la tâche M06a. Parmi les 49 élèves qui réussissent la tâche M08, 30 avaient réussi la M06a, alors que 13 élèves avaient échoué à la M06 avec un code 2 contre seulement 3 qui avaient un code 3. Il reste également 20 élèves qui avaient réussi la M06a qui ne parviennent pas à réaliser la M08. Il faut signaler que la consigne de cette tâche est une des plus difficiles de l'EA ; elle a été scindée en trois parties apparaissant successivement, mais elle pourrait nous renseigner sur certaines difficultés d'élèves à s'approprier des consignes trop complexes.

La **tâche M09** demandait aux élèves de déposer chaque sac sur une chaise indiquant en toutes lettres le nombre de billes contenues dans le sac. Cette question a été écartée pour plusieurs raisons : d'abord, certaines erreurs de captage d'un sac sur une chaise pouvaient se produire et poser problème à des élèves qui ne remarquaient pas forcément l'erreur ; dans pareil cas, il nous a parfois été possible, en analysant les séquences de manipulation, de récupérer l'information mais pas dans ce cas, les liens logiques entre les nombres de chaque chaise n'existant pas, les élèves ne portaient ensuite leur regard que sur les chaises restantes ; enfin, les quelques élèves échouant cette tâche n'avaient pas forcément de problème de manipulation ni tous de graves problèmes dans la réalisation des autres tâches.

La **tâche M11** a présenté deux difficultés inattendues : la première est celle de compréhension de ce qu'est un podium, la seconde est l'interprétation que certains élèves ont faite de l'expression (...) *Place les trois premiers sur le podium* qui considéraient alors les trois premiers sacs, soit ceux alignés à partir de la gauche. Cette tâche, peu difficile *a priori*, nous démontre tout l'impact d'un contexte culturel et de l'implicite d'une consigne lié à ce contexte. S'assurer une bonne compréhension des situations présentées aux élèves est un enjeu majeur de la construction de tâches complexes et nous mesurons d'autant plus cette difficulté pour une tâche spécifique.

En plus des réponses à ces tâches, nous avons rajouté certaines variables supplémentaires aux tâches essentiellement porteuses soit des manipulations soit de catégories d'erreur :

- M02a_63 et M02a_112 : choix d'au moins un sac <63 , respectivement >112 ; ces variables doivent nous permettre de voir dans quelle mesure la grandeur des nombres joue un rôle dans les erreurs commises.
- M04PAR10 et M04MANIP : associées à la tâche M04 demandant de dénombrer des billes, la première fournit une information relative à l'organisation des billes par paquets de 10, la seconde donne trois catégories correspondant à un intervalle d'un nombre de manipulations; M04aPAR10 a également été construite pour M04a.
- M06MANIP : associée à M06, elle fournit des catégories de manipulation des boîtes et des billes; ainsi, pour compter des billes, certains élèves commençaient par descendre les billes une à une dans l'encadré prévu à cet effet, sans considérer les boîtes; or, seules 10 billes étaient disponibles et cette manipulation est ainsi repérable; de même lorsque aucune bille ou aucune boîte n'a été touchée, il nous a semblé qu'il pouvait être intéressant de conserver cette information.
- M08T3 : la tâche M08 était décomposée en plusieurs étapes de lecture de la consigne, si bien que nous avons tenté de construire une variable de temps relative à la lecture de cette consigne; quatre catégories ont ainsi été créées, par échelle linéaire d'une minute.
- M10T : là aussi, nous avons déterminé trois catégories de temps de réalisation de la tâche sur des écarts temporels beaucoup plus faibles (16 sec. – 10 sec. – au-delà de 28 sec.), ayant observé que cette tâche, relativement classique, était très vite réalisée; il s'agira de voir si elle a effectivement un rôle à jouer dans notre analyse.

Enfin, des variables plus globales ont été élaborées en tenant compte parfois de la combinaison de plusieurs variables dans l'EA :

- MTL : temps de lecture de la partie introductive.
- MTT : catégories de temps moyen par item obligatoire (moyenne des temps).
- MZAP : cette variable relève d'informations spécifiques à chaque tâche (manipulations, temps de réalisation, souhait de passer à la tâche suivante,...) permettant d'observer le comportement des élèves sur les tâches à réaliser.
- MPRO : cette variable est basée sur quatre tâches (M01, M06, M08, M11) relevant de compétences de degré situé entre 2 et 3 selon Rey et al. (2004); il s'agit en effet de problèmes mathématiques qui peuvent paraître des problèmes d'application mais qui ne le sont vraisemblablement pas pour tous les élèves. Nous avons repéré les manipulations et la démarche de résolution sur ces quatre tâches, cette variable tentant donc d'estimer la façon dont les élèves entraient dans une résolution de problème, sans toutefois qu'il s'agisse réellement de tâches complexes selon nous; les degrés de compétences n'étant pas absolus, il nous paraissait intéressant d'observer dans quelle mesure certaines tâches *a priori* non complexes pouvaient toutefois l'être pour certains élèves.

- MREU: variable construite sur le cumul des réussites aux tâches, elle sert avant tout à repérer les élèves réussissant facilement, ou au contraire très difficilement.

Comme pour l'épreuve de français, nous disposons des informations relatives à l'âge des élèves, au genre ainsi qu'à la langue maternelle.

Analyse des correspondances multiples

Pour l'EA, les graphiques issus de cette analyse, nous ont permis de faire plusieurs hypothèses, orientant notre regard d'abord vers l'idée de déterminer des « profils de difficultés des élèves ».

Le premier plan factoriel (voir figure 5.2.3a) nous permet un premier constat rapide : globalement, les catégories de réponses correspondant à une bonne réponse semblent se situer dans la zone correspondant à $x < 0$ et $y > 0$ (à gauche, en haut). Pour rappel, le but n'est pas, dans une ACM, de définir les axes mais d'analyser la proximité ou l'éloignement des variables et des catégories ainsi dispersées.

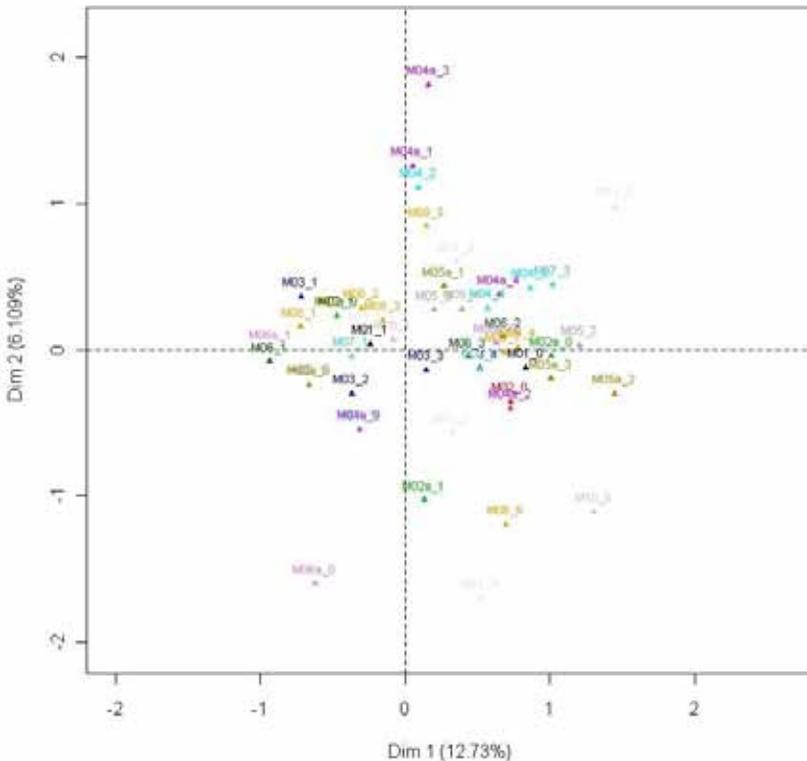


Figure 5.2.3a : Répartition des catégories de variables principales pour l'EA mathématiques

Toute catégorie de non-réponse (code 9) à une tâche complémentaire (repérable avec un a, donc de forme Mxxa_9) se superpose à celle correspondant à la réussite à l'item précédent (Mxx_1); par exemple, M05_1 est superposé à M05a_9, les élèves ayant

réussi la tâche M05 ne passant pas la M05a. Certaines catégories étant très faiblement représentées, leur éloignement de l'origine grandit rapidement, si bien que leur position est moins interprétable et que nous ne pouvons guère en tenir compte. Par exemple, il est possible d'interpréter la catégorie M04_2 (comptage peu précis du nombre de billes, 27 élèves) proche de la catégorie M04a_1 (deuxième comptage, juste cette fois; 21 élèves, dont 16 ayant eu M04_2) comme: *l'élève qui n'est pas assez précis et à qui on dit qu'il n'a pas la bonne solution semble parvenir à recompter plus fiablement cette fois*; toutefois, la catégorie M04a_3 placée un peu plus haut (ne comportant que 4 élèves), semble située assez près des deux autres (M04_2 et M04a_1); si cette zone relevait *a priori* d'élèves pouvant recommencer une tâche et la faire juste alors, cette catégorie ne pourrait être située à cet endroit³⁶. Toutefois, la catégorie M08_5 (13 élèves) est relativement proche elle aussi de nos deux premières catégories évoquées ci-dessus; M08_5 correspond à la réponse 68 lorsque la question est de connaître combien de lignes de 10 billes l'on peut former avec 68 billes, réponse que l'on interprète comme *l'élève semble perdre de vue la question* (avec des temps de 1 min. à 4,5 min.). En conclusion, cette zone ne nous a pas paru interprétable telle quelle et nous n'en avons tiré aucune hypothèse.

Nous sommes rapidement passés à l'analyse du même plan sur lequel les catégories des variables complémentaires ont été projetées (voir figure 5.2.3b).

³⁶ En fait, cette catégorie est probablement située là car elle contient 3 élèves ayant eu le code M04_2 parmi les 4 élèves concernés; c'est là la limite de l'utilisation d'une ACM avec des catégories trop faiblement représentées.

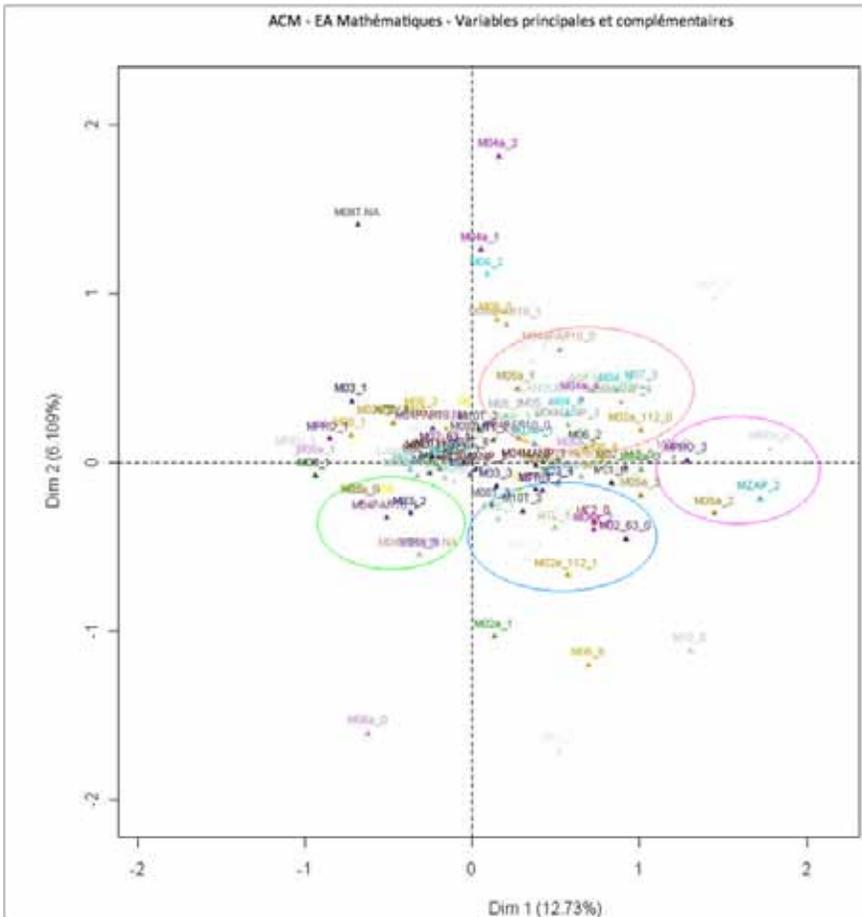


Figure 5.2.3.b: Répartition des catégories de variables principales et complémentaires pour l'EA mathématiques.

Nous nous référerons aux quadrants I ($x > 0, y > 0$), II ($x > 0, y < 0$), III ($x < 0, y < 0$) et IV ($x < 0, y > 0$)

Conscients qu'il est difficile d'une part de rendre compte de tous les regards et analyses détaillées de ce plan et, d'autre part, de le lire sans disposer des coordonnées de chaque catégorie, nous nous contentons ici de livrer globalement les constats auxquels nous arrivons. Nous avons repéré quatre zones qui nous semblent prendre sens dans une interprétation d'éventuelles difficultés des élèves; ces zones sont représentées approximativement par les formes ovales dans la figure 5.2.3b, mettant de côté la partie correspondant aux élèves qui réussissent bien.

- a) **La zone en rose**: elle contient d'abord la catégorie MZAP_2 qui est la variable construite à travers l'EA et qui correspond à des élèves tentant de passer rapidement à la tâche suivante ou ne passant que si peu de temps sur une tâche qu'ils ne sauraient y entrer. Il est intéressant de constater qu'elle se situe à proximité de la catégorie M05a_2, qui dénote, pour une tâche complémentaire, une réponse

pour laquelle l'élève n'a visiblement pas pris en considération la consigne et pour laquelle il a prélevé une information un peu à la volée (certains prenant 30 sec., d'autres 2 min.). La catégorie MPRO_3, correspondant à une difficulté de celui-ci à rentrer dans des situations (variable construite sur la base des quatre tâches assimilables à des problèmes), est également à proximité. Selon nous, les élèves dans cette zone ne rentrent simplement pas dans les tâches qui leur sont demandées; une remédiation autre que mathématique serait d'abord nécessaire pour eux.

- b) **La zone en rouge**: les catégories situées à cet endroit dénotent un échec assez massif aux tâches principales comme complémentaires; c'est notamment le cas des catégories M04_4 et M04a_4, M05_4 (voire même M05_3), M07_3, M06MA-NIP_4, MZAP_1, ... MPRO_3 est à la limite entre cette zone et la zone en rose. *A priori*, même si quelques-uns de ces élèves ont une petite tendance à vouloir passer certaines tâches, la plupart tentent de résoudre les tâches demandées mais semblent ne pas disposer de structuration des nombres à laquelle se raccrocher. Un travail conséquent serait nécessaire pour ces élèves qui commencent le cycle 2 sans savoir dénombrer une quantité jusqu'à 47, pire, sans repérer de décomposition liée aux dizaines; nous considérons qu'une reconstruction profonde de la compréhension du nombre pourrait s'avérer nécessaire chez ces élèves.
- c) **La zone en bleu**: les catégories situées là relèvent *a priori* d'erreurs souvent considérées comme de l'inattention; ainsi se trouvent M02_0, M02_63_0, M02a_112_1, M04a_2, MTL_4, MPRO_2, AGE_1, MTT_4, M07_2, ... Ces catégories indiquent des erreurs plutôt sur des nombres au-delà de 100, les élèves concernés semblent jeunes, prennent du temps et peinent encore à bien appréhender les « petits problèmes »; ils sont toutefois capables de comprendre ce qu'il y a à faire, manquant visiblement d'efficacité dans la manipulation des outils.
- d) **La zone en vert**: les catégories situées ici ne seraient pas forcément à mettre en évidence, les élèves réussissant relativement bien l'épreuve dans son ensemble. Les catégories concernées sont M04_1, M04PAR10_1, M05_1, M03_2, M06_1 (et plus proche de l'origine, M08T_2, M10T_1, MTT_1, MTT_3 et MZAP_0). Ces élèves semblent appliquer correctement les procédures apprises, font preuve de rigueur et de persévérance; alors pourquoi les mettre en évidence dans la perspective d'une visée diagnostique? Le groupe de travail a considéré ces élèves comme présentant un risque d'être en difficulté dans la résolution d'une tâche complexe, notamment parce que la catégorie MPRO_1 se situe dans le quadrant du dessus; ils sont, potentiellement, ce que Roegiers (2004) appelle des « faux forts » élèves (p. 155). Nous verrons, lors du croisement des résultats entre l'EA et l'EP si cette hypothèse est corroborée.

Par l'analyse de ces zones, nous avons établi quelques hypothèses sur des remédiations possibles quand nous constatons des déficiences relatives soit à l'investissement des élèves dans la tâche, soit à des ressources qui paraissent manquer, soit encore à la mise en œuvre des démarches.

Relevons au passage que des tâches extrêmement proches comme M03 et M07 (toutes deux concernant des nombres d'une suite obtenue par comptage de 10 en 10 à partir d'un nombre donné, l'action demandée à l'élève étant toutefois différente), que nous aurions tendance à considérer comme similaires ou relevant au moins de la même compétence, montrent des répartitions très différentes sur le plan: ainsi, réussir M03 semble plus difficile (M03_1 dans le quadrant IV) que M07 (M07_1 est

avec M03_2 (juste mais manque quelques réponses) dans le quadrant III); ceci peut toutefois trouver une logique dans le fait que l'ordre établi dans M07 peut faciliter le travail. Les catégories M03_3, M03_4 et M07_2 se situent toutes dans le quadrant II, relevant quelques erreurs, quand M07_3 se situe dans le quadrant I. Cette répartition suit relativement bien les zones caractérisées plus haut.

L'ÉPREUVE PAPIER (EP)

Afin de s'assurer que tous les élèves aient le temps d'effectuer l'ensemble de l'EP et pour ne pas alourdir la passation demandée aux enseignants volontaires, le nombre de tâches complexes a été restreint pour l'épreuve papier. On a constaté que certains élèves terminaient l'EP en moins de 10 minutes quand d'autres prenaient plus de 40 minutes. Toutefois, le lien entre les démarches investies par les élèves et le temps passé à la réalisation n'est pas direct; l'idée que le temps est un critère pertinent pour prédire la réussite ou l'échec à la résolution de la tâche doit donc être écartée.

Comme pour les épreuves de français, le groupe a effectué le codage des réponses, réajustant les codes et les catégories en conséquence. En mathématiques, le nombre d'informations prélevées dans chacune des tâches complexes reste assez faible, surtout en comparaison de toutes celles obtenues dans l'EA. Les variables nous permettant d'observer dans quelle mesure les élèves utilisent les zones d'essai mises à leur disposition n'ont guère apporté d'informations, en particulier lorsque les élèves ont laissé ces espaces vides. Nous faisons l'hypothèse que les enseignants n'incitent pas tous de la même manière leurs élèves à explorer les solutions et les procédures possibles dans la résolution de problèmes et donc, que ces derniers ne sont pas également préparés à faire des essais et à les écrire.

Variables prélevées

Variables portant sur les caractéristiques de l'élève: le genre, l'âge et la langue parlée à la maison.

Variables-réponses:

1. Pour la tâche *Le spectacle*, la réponse a été conservée telle quelle (mais peu traitable statistiquement) et simplement codée en une variable dichotomique (MP1a). Les nombres inscrits sur les sièges par les élèves ont permis de constituer une variable dichotomique (MP1b) afin d'observer si seule la partie gauche des sièges avait été investie ou non. Selon les nombres inscrits sur les sièges, une variable multi-catégorielle identifiant le type de comptage effectué par les élèves a été constituée (MP1c). Enfin, l'espace réservé sur la feuille permettant de recueillir les explications des élèves comportait suffisamment d'explications (également sur rappel de l'enseignant), nous avons pu créer une variable multi-catégorielle (MP1d) observant la cohérence entre la démarche effectuée et l'explication donnée: relevons que certains élèves découvrent alors la possibilité de procéder par un comptage de 10 en 10 ce qui a fait l'objet d'une catégorie à part entière bien qu'ils ne soient que 4 élèves dans ce cas.

2. La tâche *La tombola* comporte trois parties :

- a) La première partie, d'abord conçue pour permettre une bonne appropriation de la situation, demande aux élèves de choisir simplement des nombres pour respecter un ordre croissant de nombres. Alors que nous avons prélevé le type d'erreur (voir annexe 8.4.4), le fait que si peu d'élèves échouent rend la distinction de ces erreurs non exploitable et a donc conduit le groupe de travail à ne conserver qu'une variable dichotomique (MP2a1), traçant uniquement la réussite ou l'échec. Une autre variable dichotomique (MP2a2) permet de savoir si les élèves ont effectué des essais.
- b) La deuxième question permet d'observer dans quelle mesure les élèves respectent les contraintes dans l'utilisation des chiffres donnés pour constituer les nombres recherchés, deux solutions étant possibles ; une variable multi-catégorielle (MP2b1) retrace les erreurs des élèves et une variable dichotomique (MP2b2) permet de savoir si les élèves ont effectué des essais.
- c) La troisième question permet, comme pour la question b), d'observer la prise en compte des contraintes relatives aux nombres ; deux variables multi-catégorielles sont définies. La première (MP2c1) concerne le respect des chiffres utilisés, la seconde (MP2c2) concerne le respect de l'ordre des nombres. Enfin, une variable dichotomique (MP2c3) permet, comme pour les questions a) et b) de retracer les essais effectués.

Dans les trois questions, les zones d'essais n'ont pas été exploitées comme nous l'avions imaginé ; en effet, certains élèves effectuent des essais mais directement dans la zone destinée à la réponse finale, bien que la gomme n'ait pas été autorisée³⁷.

Premiers résultats

Tâche 1 – *Le spectacle*

Tableau 5.2.3c : *Le spectacle – Répartition des réponses en pourcents*

%	MP1a	MP1b	MP1c*	MP1d**
Code 3			31%	
Code 2			32%	3%
Code 1	67%	30%	24%	80%
Code 0	31%	37%	3%	5%
Code 9	2%	33%	9%	13%

- * Code 3 = comptage de 10 en 10 ;
 Code 2 = comptage de 2 en 2 ;
 Code 1 = comptage de 1 en 1 ;
 Code 0 = comptage non-identifiable ;
 Code 9 = aucun comptage effectué

- ** Code 2 = découverte de l'organisation de 10 en 10 ;
 Code 1 = explication cohérente avec la démarche ;

³⁷ Cette contrainte semble avoir parfois gêné les élèves ; cela peut s'expliquer, d'une part par le fait que – même adulte – on procède à des essais d'abord désordonnés (pour voir !) souvent sur la première zone disponible (en l'occurrence celle destinée à la réponse), et d'autre part par le fait que l'habitude de laisser trace de sa pensée n'est pas toujours inscrite dans les pratiques de classe.

Code 0 = incohérence entre explication et démarche;
Code 9 = aucune explication n'est donnée

Si cette tâche ne semble pas extrêmement complexe *a priori* pour des élèves de 8-9 ans, il est intéressant de constater qu'elle nécessite le recours à l'organisation des nombres et qu'elle permet de révéler très clairement leur démarche. Le taux de réussite n'est pas aussi élevé que nous³⁸ l'aurions cru mais très peu d'élèves ne sont pas du tout entrés dans la tâche et ils sont nombreux à faire des essais. Comme on peut le lire dans le tableau 5.2.3c1, sur les 42 élèves qui n'ont écrit aucun nombre sur les sièges, 34 ont trouvé la bonne réponse (soit par des démarches expertes soit par comptage de tête) quand seulement 7 élèves ne l'ont pas trouvée; ces valeurs correspondent à 40% des élèves parmi ceux qui trouvent la bonne réponse contre 18% parmi ceux qui ne la trouvent pas.

Tableau 5.2.3c1 : Le spectacle – Croisement de l'exploration des nombres (MP1b) et de la réponse (MP1a)

		Réponse – MP1a				
		1	0	9		
Nombres inscrits – MP1b	1	23 18%	15 12%	0 0%	38	30%
	0	28 22%	17 13%	2 2%	47	37%
	9	34 27%	7 6%	1 1%	42	33%
		85	39	3	127	
		67%	31%	2%		

Répartition dans chaque catégorie de réponses				
		Réponse – MP1a		
		1	0	9
Nombres inscrits – MP1b	1	27%	38%	0%
	0	33%	44%	67%
	9	40%	18%	33%
		100%	100%	100%

Bien que le calcul d'une répartition en pourcents dans ces catégories atteigne sa limite, on constate une répartition relativement similaire entre les élèves qui investissent tous les sièges ou ceux qui n'investissent que les sièges de gauche.

Répartition selon les nombres inscrits					
		Réponse – MP1a			
		1	0	9	
Nombres inscrits – MP1b	1	61%	39%	0%	100%
	0	60%	36%	4%	100%
	9	81%	17%	2%	100%

Étonnamment, la répartition selon les parties investies montre une répartition similaire de la réussite: que l'élève inscrive des nombres uniquement sur les sièges de gauche ou également de droite, ou qu'il procède par comptage de 2 en 2 ou de 1 en 1.

27% des élèves, très sûrs de leur comptage ou de leur repérage dans l'organisation des nombres, peuvent ne rien inscrire sur les sièges et trouver la bonne réponse (catégorie MP1a=1 et MP1b=9). Ceci est rassurant dans la mesure où, en fin de cycle 1, il

³⁸ Y compris les enseignants estimant la difficulté de la tâche avant passation.

est attendu que les élèves comprennent bien l'organisation des nombres jusqu'à 200 afin qu'ils puissent étendre le champ numérique à des grands nombres. Toutefois, il reste encore beaucoup d'élèves (13%) qui ont besoin de noter tous les nombres (ou presque) et qui ne trouvent pas la bonne réponse (Catégorie MP1a=0 et MP1b=0). Ceci montre une faible fiabilité dans le comptage qui prêterite les élèves déjà en instabilité sur l'organisation décimale des nombres <200 et ce d'autant plus dans la démarche exploratoire qu'ils doivent de fait lancer.

Nous posons l'hypothèse que les élèves qui recouraient à une démarche experte de résolution (soit comptage de 10 en 10) réussissaient mieux que les autres; ceci correspondrait également à constater que les élèves qui ne réussissent pas recourent à des démarches très exploratoires et n'exploitent pas l'organisation décimale des nombres.

Tableau 5.2.3c2 : Le spectacle – Croisement de la réponse (MP1a) et de la démarche (MP1c)

		Démarche – MP1c						
		3	2	1	0	9		
Réussite – MP1a	1	32	26	18	3	6	85	67%
	0	8	13	13	1	4	39	31%
	9	0	2	0	0	1	3	2%
		40	41	31	4	11	127	
		31%	32%	24%	3%	9%		

		Démarche – MP1c				
		3	2	1	0	9
Réussite – MP1a	1	80%	63%	58%	75%	55%
	0	20%	32%	42%	25%	36%
	9	0%	5%	0%	0%	9%
		100%	100%	100%	100%	100%

		Démarche – MP1c					
		3	2	1	0	9	
Réussite – MP1a	1	38%	31%	21%	4%	7%	100%
	0	21%	33%	33%	3%	10%	100%
	9	0%	67%	0%	0%	33%	100%

Le constat est rapide: les élèves recourant effectivement à un comptage de 10 en 10 (code 3 pour MP1c) réussissent dans 80% des cas alors qu'elle n'est plus que de 63% pour un comptage de 2 en 2 (code 2 pour MP1c) et à peine moins pour un comptage de 1 en 1. Les 5% d'élèves ne fournissant pas de réponse alors qu'ils ont organisé un comptage de 2 en 2 peuvent se rajouter aux 32% d'élèves qui livrent une réponse fausse. Ceci tend à rapprocher la répartition entre les deux types de démarches exploratoires (2 en 2 et 1 en 1), permettant de supposer qu'elles sont toutes deux relativement similaires en efficacité.

3% des élèves qui donnent une réponse erronée semblent faire des essais sans les organiser et 10% fournissent une réponse sans laisser trace d'éventuels essais. Le constat fait précédemment que des élèves parviennent à la réponse sans nécessairement inscrire des nombres, ou encore que certains élèves ne parviennent pas à la bonne réponse après avoir bien organisé un comptage (de 2 en 2 par exemple), semble montrer qu'aucune démarche n'est absolument efficace, mais que la réussite à cette tâche relève d'abord d'une compréhension dans l'organisation des nombres et du repérage de cette organisation dans la situation présentée.

Globalement, cette tâche nous paraît être une situation particulièrement adaptée à des élèves de cet âge et bien ajustée au repérage du degré de compétence relatif à l'organisation décimale des nombres.

Tâche 2 – La tombola

Tableau 5.2.3d : La tombola – Répartition des réponses en pourcents

%	MP2a1	MP2a2	MP2b1*	MP2b2	MP2c1**	MP2c2***	MP2c3
Code 11			28%		21%	91%	
Code 10			32%		3%	0%	
Code 1	95%	31%	31%	31%	39%	6%	43%
Code 0	5%	23%	6%	13%	35%	2%	17%
Code 9	0%	46%	2%	56%	2%	2%	40%

- * Code 11 = deux réponses justes;
 Code 10 = au moins une des deux réponses est cohérente;
 Code 1 = encadrement correct mais utilisation d'autres chiffres;
 Code 0 = mauvais encadrement et utilisation d'autres chiffres;
 Code 9 = aucune réponse
- ** Code 11 = bon choix et respect des chiffres;
 Code 10 = bon choix des chiffres pour un nombre correct > 89, mais erreur dans les nombres < 89;
 Code 1 = bon choix des chiffres pour les nombres corrects < 89 mais erreur pour nombres > 89;
 Code 0 = utilisation d'autres chiffres dans les nombres de part et d'autre de 89;
 Code 9 = aucune réponse
- *** Code 11 = ordre des nombres respecté;
 Code 10 = nombre ordonné >89;
 Code 1 = nombres ordonnés <89;
 Code 0 = ordre des nombres non respecté;
 Code 9 = aucune réponse

Comme nous l'avons décrit un peu plus haut, la première partie de la tâche (a) doit d'abord rassurer les élèves afin qu'ils comprennent la situation (tout au moins qu'ils l'associent à l'activité préalable) et ils ont à choisir des nombres compris entre deux autres. Cette partie a) a l'avantage de nous prouver, par ses résultats très nets, que la majorité des élèves (95%) comprend parfaitement la situation ainsi que la question mathématique sous-jacente. Toutefois, les 5% restant (catégorie MP2a1=0) sont à observer de près dans notre épreuve diagnostique; nous y reviendrons.

La partie b) nécessite l'utilisation de trois chiffres pour construire deux nombres, l'un plus petit que 73 et l'autre plus grand. Seulement 28% des élèves y parviennent, alors que 32% semblent ne parvenir à réaliser qu'une étape de la démarche, soit à composer un seul des nombres recherchés; en effet, il est nécessaire de revenir sur son premier choix si jamais le second nombre ne convient pas. 31% des élèves restent bloqués sur la première contrainte de respecter l'ordre des nombres, mais ne respectent pas la nouvelle contrainte liée aux chiffres à disposition. Plus grave encore, 6% des élèves ne respectent ni l'ordre des nombres ni la restriction du choix de chiffres et 2% ne répondent pas du tout.

La partie c), relativement similaire à la partie b), mais complexifiée par l'augmentation du nombre de chiffres à disposition et par la nécessité de constituer un nombre à trois chiffres, impliquant qu'il y en ait aussi un à un seul chiffre. Ils ne sont alors plus que 21% à trouver des nombres valides (catégorie MP2c1=11). Seulement 3% des

élèves (catégorie MP2c1 = 10) parviennent à constituer un nombre plus grand que 89 en respectant le choix des chiffres, mais pas pour les nombres plus petits que 89 ; par contre 39% respectent le choix des chiffres (catégorie MP2c1 = 1) pour constituer les nombres plus petits que 89 sans y parvenir pour un nombre plus grand que 89. Ils sont 35% au demeurant à ne plus respecter aucune contrainte (catégorie MP2c1 = 0), y compris celle de départ posée dans a) de respecter un ordre croissant des nombres obtenus ; le fait de changer un chiffre semble pour ces élèves devoir immanquablement faire perdre toute cohérence dans l'ordre des nombres, sans qu'il soit possible d'agir sur cet ordre selon le chiffre choisi.

Bien que les élèves soient jeunes, nous constatons à quel point il leur est difficile de revenir sur leur résultat. Dans cette tâche, aucune demande de l'enseignant de fournir une explication n'a été formulée ; une telle sollicitation pourrait-elle conduire à une meilleure réussite ? Il est toutefois difficile pour la plupart des élèves de revenir sur leur production ; ils préfèrent parfois trouver une explication qui colle au résultat sans forcément relater la réelle démarche mise en œuvre.

Les tâches b) et c) étant très proches, nous nous sommes demandé si un effet existait entre les deux, appuyant ou non le choix de les avoir proposées successivement. Le tableau 5.2.3d1 montre :

- d'une part, une répartition des élèves ayant réussi la tâche 2b) dans des catégories de réussite diverses dans la tâche 2c) ;
- d'autre part, un échec assez régulier à la tâche 2c) pour les élèves n'ayant que peu réussi à la tâche 2b).

Ces deux constats démontrent effectivement une complexification de la tâche demandée dans 2c) par rapport à 2b). Nous n'aurions certainement obtenu aucune information sur ces élèves en difficulté si nous avions proposé uniquement la partie c).

Tableau 5.2.3d1 : Croisement du type de réussite entre la partie b) et c) de la tâche La tombola

		Tombola b)						
		11	10	1	0	9		
Tombola c)	11	16 13%	8 6%	3 2%	0 0%	0 0%	27	21%
	10	1 1%	2 2%	1 1%	0 0%	0 0%	4	3%
	1	16 13%	17 13%	12 9%	3 2%	1 1%	49	39%
	0	3 2%	14 11%	23 18%	5 4%	0 0%	45	35%
	9	0 0%	0 0%	1 1%	0 0%	1 1%	2	2%
		36	41	40	8	2	127	
		28%	32%	31%	6%	2%		

Dans cette tâche, nous avons fait le choix de procéder par étapes, s'approchant plutôt de ce que Rey et al. (2004) auraient associé à une compétence de degré 2. Ne présenter que la partie c) de la tâche entraînait le risque de n'obtenir que trop peu de réponses, alors que bien des élèves sont capables de rentrer dans une telle tâche. Ce choix soulève toutefois un éventail de questions quant à la définition d'une tâche complexe, associée à une compétence de degré 3 :

- Une tâche complexe peut-elle contenir plusieurs questions? Si oui, quelles sont les contraintes à respecter pour relever d'une compétence de degré 3 et pas de degré 2?
- Une succession de tâches relatives à une même situation exclut-elle qu'une d'entre elles puisse relever effectivement d'une compétence de degré 3?

Lien entre les tâches *Le spectacle* et *La tombola*

À la création de l'épreuve, nous avons pour hypothèse que les élèves qui repèrent effectivement un comptage de 10 en 10 dans la tâche *Le spectacle* devaient être au point sur l'organisation du système décimal de position et par conséquent comprendre rapidement l'enjeu dans les tâches b) et c) de *La tombola*.

Tableau 5.2.3.d2: Croisement du type de démarche entre la tâche *Le spectacle* et les parties b) et c) de la tâche *La tombola*

		Le spectacle					
		3	2	1	0	9	
<i>La tombola b)</i>	11	9%	11%	5%	0%	3%	28%
	10	9%	9%	11%	1%	2%	32%
	1	10%	9%	7%	2%	3%	31%
	0	2%	3%	0%	1%	0%	6%
	9	0%	0%	2%	0%	0%	2%
		31%	32%	24%	3%	8%	100%
<i>La tombola c)</i>	11	9%	8%	2%	1%	2%	21%
	10	0%	2%	2%	0%	0%	3%
	1	13%	9%	12%	1%	4%	39%
	0	10%	13%	9%	2%	2%	35%
	9	0%	1%	1%	0%	0%	2%
		31%	32%	24%	3%	9%	100%

À la lecture du tableau 5.2.3.d2, nous constatons un croisement très important dans les catégories de réponses. Notre hypothèse n'est pas vérifiée, au contraire: les résultats montrent que la compétence qui nous semblait être mise en œuvre dans ces tâches n'est peut-être pas la même, ou – préférons-nous penser – les tâches révéleraient chacune certaines dimensions de la compétence évaluée. Décider alors d'un niveau de réussite prend sens pour informer sur la compétence évaluée.

Concernant la forme des épreuves, si le support papier semble relativement habituel et *a priori* adapté à la réalisation de tâches complexes, deux constats sont à retenir et à exploiter pour la création d'épreuves futures :

1. Le recours à l'exploration semble être extrêmement variable d'un enseignant à l'autre ; il semble donc difficile à de si jeunes élèves d'investir des espaces destinés à des essais sans en avoir l'habitude, même si l'enseignant en rappelait l'existence lors de la présentation du travail à mener et pouvait créer le lien avec une éventuelle pratique de classe. L'habitude pourrait s'installer par le biais d'une telle épreuve, mais il serait également intéressant, lorsque des activités préalables existent, d'en soumettre aux élèves afin qu'ils puissent soit les associer à une pratique existante, soit en comprendre l'enjeu.
2. L'habitude d'une certaine présentation du travail soumis aux élèves (à l'instar du témoignage d'une enseignante) doit interroger d'une part l'uniformité des documents soumis en classe, d'autre part celle des épreuves à constituer. Varier les présentations en classe produit un apprentissage, certes souvent considéré comme secondaire par rapport à celui visé dans les disciplines. S'agirait-il d'en privilégier la pratique et si oui, à partir de quel âge ?

Faut-il agir sur les pratiques ? La possibilité de présenter des épreuves « à blanc » (voire des activités préalables aux épreuves) ne suffirait-elle pas à gommer l'effet d'*habitus* lié au support et au travail en classe ? Ces questions se rapprochent du constat fait sur la simplification systématique des consignes par A. Guay (2010).

ÉVALUATION EN DEUX PHASES (EA+EP)

Dans le but de procéder à un rapide constat des réussites croisées entre l'EA et l'EP, nous avons constitué un score pour chaque épreuve en nous basant non seulement sur les réponses, mais également sur les procédures mises en œuvre, surtout visibles dans l'EP. Pour l'EA, nous avons simplement constitué un score basé sur la réussite aux tâches obligatoires (donc effectuées par tous les élèves) et en avons défini quatre catégories (code 4 = entre 0 et 3 tâches réussies, code 3 = entre 4 et 5 tâches réussies, code 2 = entre 6 et 7 tâches réussies, code 1 = plus de 8 tâches réussies). Pour l'EP, nous avons attribué un maximum de 2 points pour la tâche *Le spectacle*³⁹, pour la tâche b) de *La tombola* ainsi que pour la tâche c) ; ensuite, nous avons créé trois catégories (code 3 = entre 0 et 1 point, code 2 = entre 2 et 3 points, code 1 = entre 4 et 6 points (soit ici au moins une tâche à 2 points)).

Nous avons enfin procédé à un croisement entre ces catégories. Nous cherchions à vérifier l'hypothèse souvent entendue qu'il faut d'abord maîtriser les outils pour réussir des tâches plus complexes ; dans notre cas, nous devrions constater une bonne réussite à l'EP seulement pour les élèves réussissant bien l'EA.

³⁹ Pour exemple de prise en compte des variables, voici la qualification des points : 2 pt si MP1a=1 et MP1c=3, 0,9 (bénéfice du doute) ; 1 pt si MP1a=1 et MP1c=1, 2 ; 1 pt si MP1a=0 et MP1c=3 ; 0 pt si MP1a=0 et MP1c=0,1,2,9 ; 9 si MP2b1=9.

Tableau 5.2.3e: Croisement de la réussite de l'EA et de l'EP

		Réussite à l'EA (MREU)					
		1	2	3	4		
Réussite à l'EP (MPREU)	1	22 18%	14 11%	3 2%	3 2%	42	33%
	2	10 8%	23 19%	16 13%	4 3%	53	42%
	3	7 6%	8 6%	10 8%	4 3%	29	23%
		39	45	29	11	124	
		31%	35%	23%	9%		

D'après ces résultats, on peut conclure qu'entre 20% et 30% environ des élèves sont performants voire très performants dans l'EA et également dans l'EP, et qu'ils sont à peu près autant à l'être un peu moins sur des tâches complexes. Les autres élèves, soit environ 40% présentent certaines difficultés qu'il s'agit d'identifier pour évaluer la nécessité – ou non – de remédiation :

<p>Élèves réussissant l'EA mais pas l'EP (12%) (voir tableau 5.2.3e)</p>	<p>Ces élèves seraient qualifiés par Roegiers (2004) de <i>faux forts</i> élèves (p.155); en effet, ils semblent à l'aise sur des tâches spécifiques, le caractère informatisé de ces tâches ne leur pose pas de problème et on peut donc considérer qu'ils sont capables de transposer leurs connaissances sur des tâches qu'ils (re)connaissent. Toutefois, les tâches complexes leur posent des difficultés et leurs connaissances, bien que très au point pour la moitié d'entre eux, ne leur permettent pas de réussir complètement ne serait-ce qu'une des tâches complexes proposées, même dans le cas de la tâche <i>Le spectacle</i> où ils procèdent par exploration.</p>
<p>Élèves réussissant l'EP mais pas l'EA (4%) (voir tableau 5.2.3e)</p>	<p>Ils sont beaucoup moins nombreux dans ce cas, constat tant soit peu rassurant sur le rôle de l'école dans les apprentissages scolaires. Toutefois, bien que fort peu nombreux, ces élèves font preuve d'une réussite remarquable sur les tâches complexes au regard de leur faible réussite sur des tâches spécifiques. Roegiers (2004) les qualifierait de <i>faux faibles</i> élèves (p. 157) car ils présentent visiblement de bonnes compétences dans la compréhension du système décimal mais ont des difficultés dans des tâches plus formelles.</p>
<p>Élèves ne réussissant que peu ou très peu à l'EA et à l'EP (14%) (voir tableau 5.2.3e)</p>	<p>Ces élèves sont encore nombreux alors qu'ils sont en début de cycle 2 et donc promus en fin de 4^e. Parmi eux se trouvent des élèves repérés comme démotivés (cf. résultats de l'EA) mais bien d'autres présentent ce profil de résultats et semblent pourtant s'investir dans le travail qui leur est demandé.</p>

Bien des hypothèses extérieures aux épreuves elles-mêmes pourraient expliquer de tels résultats et il s'agirait très certainement d'analyser d'autres paramètres externes. Cependant, l'identification de ces trois types de profils présente l'intérêt d'ouvrir l'exploration de remédiations ciblées, voire de considérations sur la succession des apprentissages.

Comme nous l'avons évoqué dans l'analyse de l'EP, et parce que nous travaillons dans une visée diagnostique, revenons sur les 5% d'élèves repérés (voir tableau 5.2.3d) qui ne sont pas parvenus à choisir des nombres pour des billets de *La tombola* afin de les placer dans un ordre croissant. En observant de près ces élèves (6), on constate qu'ils se répartissent autant en bonne réussite sur la tâche *Le spectacle* et sur la partie b de *La tombola* qu'en échec voire non-réponse. En outre, leur résultat à l'EA se répartit également entre bonne réussite et échec. Le fait que l'EA comme l'EP en mathématiques ne comporte aucun QCM à réponse unique, la part de chance dans les scores obtenus ne peut être tenue pour responsable de la réussite (ou de l'échec) de ces élèves. Bien que leur nombre soit réellement réduit (ce qui est heureux dans une telle épreuve), on peut considérer qu'il existe, comme nous l'avons constaté grâce au tableau 5.2.3e, des élèves capables de réussir des tâches spécifiques mais pas des tâches complexes, l'inverse étant également vrai ! Ces élèves interrogent : la disparité dans la réussite aux tâches rend difficile toute prédiction de succès. Il serait certainement utile pour ces élèves – mais aussi pour la compréhension que peut s'en faire l'enseignant – de leur demander une justification systématique de leur choix dans toute réalisation.

En conclusion

Contrairement à l'évaluation en deux phases en français, celle en mathématiques, tout en portant de façon très ciblée sur les attentes fondamentales en lien avec l'objectif d'apprentissage Poser et résoudre des problèmes pour construire et structurer des représentations des nombres naturels s'est concentrée surtout sur la mesure de la compréhension par les élèves du système décimal de position. Il a en effet paru essentiel au groupe de travail de vérifier cette compétence, centrale dans l'appréhension des nombres, et qui semble source, ultérieurement et chez de nombreux élèves, de difficultés relatives aux nombres réels dans leur ensemble.

D'après nos constats, il est nécessaire de vérifier finement l'état des connaissances des élèves relativement aux attentes fondamentales, ce qui nous semble avoir été parcouru dans l'EA, bien que des améliorations de certaines tâches puissent encore être apportées. Toutefois, il serait incomplet de se limiter à une mesure relative aux ressources, et il nous paraît indispensable de présenter des tâches complexes mettant en évidence la mobilisation des ressources évaluées dans l'EA et mesurant les composantes d'une résolution également considérée comme attente fondamentale.

En outre, l'observation des réalisations dans l'EP nous apprend qu'il serait nécessaire de distinguer les caractéristiques de tâches complexes d'apprentissage des tâches complexes d'évaluation. En effet, lors d'une évaluation, les réalisations des tâches complexes devraient idéalement démontrer le recours à une démarche experte ; or seul un petit tiers des élèves y recourt. Mais permettre aux élèves de rentrer dans une démarche exploratoire lors d'une évaluation atteste d'une démarche mathématique qui démontre également un apprentissage. Les tâches complexes d'apprentissage pourraient en cela se distinguer de celles d'évaluation par le fait qu'une exploration

devrait effectivement être menée et aboutir à la mise en évidence d'une démarche de résolution experte, dont les ressources nécessaires à sa réalisation sont les savoirs enseignés et donc à institutionnaliser. Il serait par ailleurs intéressant d'explorer quel devrait être le niveau d'intégration de certaines compétences permettant effectivement leur mobilisation dans des démarches expertes propices à la réalisation de tâches complexes d'évaluation.

5.2.4 Les liens entre l'épreuve français et l'épreuve mathématiques

Nous ne nous sommes pas arrêtés à une comparaison des résultats des élèves dans chacune des disciplines, mais, profitant de leur développement conjoint et dans une perspective diagnostique, nous avons choisi de développer un double questionnaire :

- le premier autour du rôle de la compréhension dans la réussite des tâches de mathématiques ;
- le second autour de critères généraux et transversaux décrivant la réussite à des tâches complexes.

Plus précisément, nous nous sommes demandé dans quelle mesure une bonne compréhension en français peut favoriser celle des consignes en mathématiques ou inversement, dans quelle mesure une mauvaise compréhension en français risque d'entraîner des problèmes de compréhension des consignes en mathématiques. En ce qui concerne les tâches complexes, nous avons construit des variables à même de nous informer sur la réussite à des tâches complexes, qu'elles soient en langue ou en mathématiques.

Exploration de la compréhension

Pour répondre à notre première préoccupation, nous avons tenté de mettre en évidence des liens entre des variables (globales ou ponctuelles) touchant à la compréhension en français et des variables en mathématiques donnant une information sur des difficultés possibles touchant à la compréhension de consignes.

En français, les variables globales retenues sont : FTL, F01, FQ7a et b et FREU ; les plus ponctuelles sont : F02, F04, F05, F06a, b et c, F09, FQ3, FQ4, FQ6, (voir point 5.2.2). Les premières touchent donc au temps de lecture du texte (celui de *Tibili le petit garçon qui ne voulait pas aller à l'école*), à sa compréhension globale, et sur le repérage de l'intention de la lettre de Paul et des indices présents dans la lettre elle-même ; le taux de réussite générale à l'EA a été également pris en considération. Les secondes concernent des questions portant sur des parties du texte de *Tibili le petit garçon qui ne voulait pas aller à l'école*, sur la segmentation de phrases en mots, sur la reconnaissance de phonèmes et de graphèmes de mots, et, dans la tâche complexe, sur le repérage des opinions de Paul et de sa maman et sur les arguments de celle-ci.

En mathématiques, nous avons d'abord retenu des variables globales fournissant des indicateurs sur la facilité de s'approprier une consigne comme MTT, MPRO, MREU, M08T. La première (MTT) est un temps moyen par item obligatoire de l'EA, mais qui n'est qu'un indice d'une rapidité dans la réalisation des tâches ; la deuxième (MPRO) est construite sur quatre tâches de l'EA (relativement proches de problèmes mathé-

matiques pour certains élèves) et représente un indice sur la façon dont les élèves s'approprient la situation; la troisième (MREU) représente un type de réussite globale sur l'EA; la quatrième (M08T) fournit un temps partiel de lecture de la consigne (présentée par étape aux élèves) fournissant un indice dans la compréhension d'une étape en particulier. D'autres variables ont été observées pour la comparaison avec la compréhension en français; il s'agit de: MP2b1, MP2c1, M05, M06MANIP, M08 (voir point 5.2.3) qui fournissent une indication sur le type d'incompréhension des élèves dans la réalisation de certaines tâches, repéré à partir du type d'erreur effectué.

Nous avons considéré les élèves ayant connu des difficultés aux tâches de français en nous demandant si leurs résultats aux tâches de mathématiques montraient le même profil. Un exemple de tableau analysé pour cette comparaison est le croisement des réussites à F01 (résumé du texte de *Tibili*) et de la variable MPRO (indice sur l'entrée dans le problème) (voir tableau 5.2.4a). Nous faisons l'hypothèse que les élèves qui réussissent à comprendre globalement un texte s'approprient plus facilement une consigne et devraient donc montrer une meilleure entrée dans le problème. On constate effectivement que les élèves trouvant le bon résumé se situent généralement dans une catégorie représentant un bon à très bon indice d'entrée dans les problèmes (34 élèves sur les 70 en catégorie 2, 26 en catégorie 1), alors qu'ils ne sont que 10 à avoir des difficultés à entrer dans les problèmes. Parmi ceux qui ne trouvent pas le bon résumé en français, ils se répartissent pratiquement autant entre les deux indices extrêmes d'entrée dans les problèmes: 9 élèves semblent très bien y entrer pour 7 qui n'y parviennent que très mal. Malgré un nombre faible d'élèves en situation d'échec sur le résumé, mais toujours dans le but de définir un diagnostic, nous avons exploré plus avant comment ces 18 élèves se répartissaient lorsqu'ils avaient à recommencer, un peu plus loin dans l'EA de français, la tâche sur le résumé (F08). Il est intéressant de constater que la plus grande partie des élèves ayant un bon indice d'entrée dans les problèmes ont tendance à trouver la bonne réponse (5 sur les 9 élèves) alors que la majorité des élèves ayant un mauvais indice d'entrée dans les problèmes refont la même erreur sur le résumé. Ces tendances mériteraient toutefois d'être corroborées par d'autres mesures et d'autres situations vu le nombre très faible d'élèves.

Tableau 5.2.4a: Croisement de la réussite à F01 et de la variable MPRO et répartition des réponses à F08

		Indice entrée dans les problèmes (MPRO)				
		1	2	3		
Réussite résumé (F01)	1	26 30%	34 39%	10 11%	70	79%
	0	9 10%	2 2%	7 8%		
		35	36	17	88	
		40%	41%	20%		
Réussite 2 ^e essai résumé (F08)*	1	5	0	0		
	2	2	1	6		
	3	2	1	1		
		9	2	7		

*Pour rappel, cette tâche est la même que F01 mais soumise à nouveau aux élèves y ayant échoué; la catégorie 1 représente la bonne réponse, la catégorie 2 représente la même erreur que la réponse à F01, la catégorie 3 représente une autre erreur que celle faite à F01.

Les valeurs obtenues ci-dessus ne peuvent pas être considérées comme résultat absolu. Elles présentent toutefois des pistes intéressantes de recherche vers un affinement du diagnostic. Nous faisons donc deux constats: d'une part les élèves qui ont de la peine dans la compréhension globale d'un texte ne semblent pas automatiquement en avoir dans les tâches de mathématiques que nous avons retenues, d'autre part ceux qui en ont dans les tâches plus ponctuelles, en particulier dans la tâche de segmentation de phrases en mots, ne sont pas tous gênés dans la réalisation des tâches de mathématiques, ni même dans l'appropriation des consignes et des situations mathématiques.

Exploration de la réalisation de tâches complexes

En nous inspirant d'une proposition de Roegiers (ADMEE 2012), nous avons constitué des variables dans le but d'effectuer une analyse comparative des tâches complexes en mathématiques et en français. Roegiers préconise de prendre en considération des critères qui puissent dire quelque chose sur la compétence des élèves à résoudre une tâche complexe, de quelque discipline qu'elle provienne. Inspiré par Vial (2001) (in Roegiers 2004) qui propose de ramener la réussite à trois grands critères que sont la justesse, la complétude et la pertinence, il propose par exemple d'en retenir trois, moins généraux et plus maniables:

- l'élève entre correctement dans la tâche (compréhension de la tâche);
- il utilise correctement les ressources (mise en œuvre des outils);
- il établit un lien correct entre la situation et la réponse donnée (adéquation).

Ces critères peuvent être affinés, l'important étant de ne pas disposer de trop de critères afin de garantir leur indépendance.

Pour notre part et dans cette perspective, nous avons, à partir de l'EP en mathématiques et de la production écrite en français, constitué trois variables pour chaque discipline:

1. FPCONS et MPCONS: compréhension de la consigne et compréhension globale de la tâche à réaliser.
2. FPREAL et MPREAL: mobilisation des bons outils pour réaliser la tâche.
3. FPADEQ et MPADDEQ: adéquation de la solution avec la situation.

Notons que ces variables ont été construites après coup, à partir des variables dont nous disposons déjà pour nos épreuves. Nous verrons ci-après que cette construction n'est de ce fait pas complètement satisfaisante, et qu'il nous semble préférable de réfléchir à ce que pourraient être ces différentes variables au moment même de l'élaboration des épreuves.

En français, pour obtenir la variable FPCONS, nous avons repris les variables suivantes: FP1, FP2, FP3, FP4. La réussite à ces variables témoigne en effet d'une bonne compréhension de la tâche demandée: le respect du format de la lettre (les élèves sont

invités à répondre à celle de Paul), l'expression de sa propre opinion, introduite par une expression modale (ce qui est demandé dans la consigne) et l'écriture d'au moins deux arguments venant appuyer l'opinion du scripteur. Cependant, le fait d'échouer à ces critères ne signifie pas forcément que la consigne ou la tâche a été mal comprise des élèves. Il se peut que ces élèves n'aient pas saisi ce qu'est une opinion et/ou un argument ou qu'ils n'aient pas su puiser dans le travail accompli lors de l'activité préalable pour trouver un contenu adéquat à leur réponse à la lettre de Paul.

La variable FPREAL est composée de critères d'évaluation portant plus spécifiquement sur les ressources permettant la mise en texte de la lettre de réponse : FP6, FP7, FP8. FP6 concerne la qualité de la syntaxe du texte, FP7 est centré sur la présence de majuscules au début des phrases et FP8 sur la correction orthographique des mots copiés à partir du réservoir de mots à la disposition de chaque élève. Contrairement à la variable précédente, qui, nous venons de le voir, ne peut fournir des indications précises sur les causes d'un échec la concernant, celle-ci permet bien de repérer des difficultés liées à la mobilisation des ressources.

Enfin, il nous a paru intéressant d'exploiter la variable FPADEQ, c'est-à-dire celle que nous avons créée au moment de l'élaboration de l'épreuve sous le nom de FPMODIF. Le lendemain de la production de texte, les élèves étaient invités à relire leur texte et, à l'aide d'un crayon de couleur, à y apporter d'éventuelles corrections en s'appuyant sur un guide de relecture qui leur était distribué. Ce guide rappelait tous les éléments demandés assurant une bonne rédaction de la lettre de réponse à Paul. Nous avons déjà constaté, à l'examen des corrections apportées par les élèves, combien cet exercice de lecture et de tentative d'amélioration du texte est difficile pour de si jeunes élèves (d'autant plus s'ils ne sont pas habitués à le faire en classe). Ils sont donc très peu nombreux à avoir apporté des corrections (N=41, soit environ 35% de la population) et les corrections portent uniquement sur des aspects formels du texte, principalement d'ordre orthographique. Nous avons repéré, pour ces 41 élèves, si les corrections apportées étaient bénéfiques ou non (ajout d'au moins une erreur) : 31 d'entre eux font une ou plusieurs corrections bénéfiques, 10, au contraire, ajoutent au moins une erreur à leur texte. Cette variable FPADEQ, dans ce cas, ne donne donc que peu d'informations sur la capacité des élèves à fournir un texte en bonne adéquation avec la situation de production. Nous pouvons faire l'hypothèse que des élèves des cycles ultérieurs apporteront des améliorations à leur texte portant à la fois sur des éléments de contenu, de planification, de mise en texte, mais aussi relatives au contexte de production. Nous serons alors mieux en mesure, par l'observation de ces corrections, de saisir celles qui révèlent la compétence des élèves à s'adapter à la situation complexe présentée, pour autant qu'elles apportent réellement une amélioration du texte en fonction des buts de la rédaction, du ou des destinataire(s) auquel il s'adresse, et, par conséquent, des consignes fournies pour la réalisation de la tâche.

En mathématiques, un travail similaire a donc été fait. Une variable MPCONS a été construite en attribuant et en cumulant des points aux catégories des variables MP1c, MP2a1, MP2b1 et MP2c1. Ces variables nous semblent en effet révéler la prise en compte des contraintes présentes dans les consignes, comme le fait d'attribuer des numéros aux sièges dans la tâche *Le spectacle* et le fait de respecter l'ordre des nombres ou le choix des chiffres donnés pour constituer ces nombres. Évidemment, il se peut que des élèves comprennent la consigne mais que les traces laissées sur le papier ne nous permettent pas d'identifier cette compréhension ; le cumul des résultats de plusieurs tâches bien choisies doit permettre d'atténuer cet effet.

La variable concernant la réalisation en mathématiques (MPREAL) reprend les variables MP1a, MP1c, MP2b1, MP2c1 en transformant là-aussi les catégories⁴⁰ de réponses en points et en les cumulant. Le degré de réalisation est en effet visible à travers certaines réponses et dans la façon dont l'élève les a construites (qu'elles soient justes ou fausses). Enfin, la variable concernant l'adéquation de la réponse (MPADEQ) a été construite sur deux autres variables: MP1d (en relevant les élèves qui, soit ont directement utilisé l'organisation par dizaines et l'ont expliqué tel quel, soit ont découvert lors de la rédaction de l'explication, l'organisation en dizaines) et MP2c2 (en relevant les élèves qui ont respecté l'ordre des nombres – sans forcément faire juste!) puisque la tâche 2a) montrait une très forte majorité d'élèves qui semblaient parfaitement capables de choisir des nombres pour assurer leur ordre croissant.

Les variables multi-catégorielles ainsi constituées pour le français comme pour les mathématiques ont trois catégories représentant un ordre croissant de point (catégorie 1 = pas/peu de points; catégorie 3 = tous/presque tous les points). En croisant les variables correspondantes entre le français et les mathématiques, nous avons tenté de repérer d'éventuels liens révélateurs d'une sorte de compétence transversale relative à la réalisation de tâches complexes. Des tableaux 5.2.4b et 5.2.4c nous ne pouvons malheureusement guère extraire une information très pertinente; la répartition dans les catégories présente des déséquilibres et notre faible nombre d'élèves ne permet plus de généralités sur leur répartition; la concentration de réponses entre la catégorie 3 en français et 2 en mathématiques peut provenir d'un effet d'accumulation dans la création des catégories comme représenter effectivement quelques pistes de recherche:

- les élèves ayant une bonne compréhension en français pourraient être avantagés pour la compréhension des tâches mathématiques, mais ceux qui ont une mauvaise compréhension ne seraient pas forcément pénalisés pour autant;
- les élèves recourant aux bonnes ressources pour la réalisation de leur production en français ne recourent par forcément aux bonnes ressources dans la réalisation de tâches complexes en mathématiques, l'inverse étant également valable.

Tableau 5.2.4b: Croisement des indices relatifs à la compréhension en français et en mathématiques

		MPCONS				
		1	2	3		
FPCONS	1	0	9	1	10	11%
	2	2	8	2	12	14%
	3	3	41	22	66	75%
		5	58	25	88	
		6%	66%	28%		

⁴⁰ N'ayant que peu de variables disponibles dans l'épreuve mathématiques, nous avons choisi de réserver certaines catégories pour MPCONS et d'autres pour MPREAL; toutefois, la dépendance existe de fait entre elles et les variables impliquées mériteraient d'être préalablement conçues indépendantes les unes des autres.

Tableau 5.2.4c: Croisement des indices relatifs à la réalisation des tâches en français et en mathématiques

		MPREAL				
		1	2	3		
FPREAL	1	1	7	4	12	14%
	2	3	9	2	14	16%
	3	11	42	9	62	70%
		15	58	15	88	
		17%	66%	17%		

L'analyse du tableau 5.2.4d présente toutefois un autre intérêt concernant l'adéquation des réponses entre le français et les mathématiques. En effet, selon le type d'adéquation de la réponse pour le français, la répartition entre les catégories pour les mathématiques semble relativement similaire (env. 1/4 pour la catégorie 1, environ 2/5 pour la catégorie 2, environ 1/3 pour la catégorie 3); inversement, la répartition dans les catégories de français est similaire en fonction de la catégorie de réponses en mathématiques (environ 2/3 pour la catégorie 1, environ 1/10 pour la catégorie 2, environ 1/4 pour la catégorie 3). Ceci révèle une indépendance assez claire entre ces deux variables à savoir que vérifier l'adéquation d'une production en français est visiblement de nature très différente que de vérifier une réponse à une tâche complexe en mathématiques.

Tableau 5.2.4d: Croisement des indices relatifs à l'adéquation de la réponse en français et en mathématiques

		MPADEQ				
		1	2	3		
FPADEQ	1	14	26	17	57	65%
	2	2	3	3	8	9%
	3	5	10	8	23	26%
		21	39	28	88	
		24%	44%	32%		

De ce croisement de résultats entre le français et les mathématiques, il nous semble difficile, en l'état de nos épreuves et des élèves testés, de produire des constats absolus. Il est vrai que la différence de complexité des tâches des EP entre celles de français et celles de mathématiques semble importante, les premières provoquant visiblement un apprentissage des élèves non négligeable dans une mesure de compétences. L'effet lié à la « proximité temporelle » d'un apprentissage n'est guère connu quant à son impact sur la mobilisation de ces nouvelles ressources en situation d'évaluation ainsi que sur leur stabilité. Toutefois, créer des variables similaires globales pour des tâches complexes d'une même discipline ou pour des disciplines différentes demeure une piste intéressante à nos yeux et mériterait des approfondissements.

5.3 Quels messages pour quels publics ?

Il ne suffit pas, à l'instar des *Livrets personnels de compétences* (LPP) proposés en France pour rendre compte du *Socle commun de connaissances et compétences*, de proposer un carnet où l'enseignant doit apposer des croix face aux dites connaissances et compétences, pour s'assurer effectivement que :

- la communication est pertinente pour les acteurs ;
- l'évaluation associée est ajustée aux objets de la mesure.

Dans son article, Rey (2012) reprend une multitude de difficultés inhérentes à la mise en œuvre du socle et du décalage existant avec le LPP. Notamment, il relève que *Le livret a été donné aux enseignants sans exercices d'accompagnement, ce qui a laissé planer le doute sur le niveau cognitif précis que les items de compétence étaient censés recouvrir* (p. 13). Qui plus est, (...) *le caractère binaire de la validation des compétences (« oui » ou « non ») semble poser problème à des enseignants (Delatouche 2012) car elle écrase les degrés de progression (...)* (p. 13) ; et le fait qu'une compétence évaluée par un enseignant peut être signalée (...) *acquise ou non acquise à la fin du primaire en fonction d'appréciations subjectives sur le niveau qu'il [l'enseignant] estime « normal » d'avoir atteint (...)* (p. 13), montre la nécessité d'une définition commune minimale des caractéristiques d'une évaluation de compétences.

Il est essentiel de tenir compte, d'une part, des caractéristiques de l'évaluation et des données qu'elle peut fournir, et, d'autre part, du destinataire du message. Ceci implique que les informations sont d'abord à construire en fonction de ce dernier et qu'il doit être décidé au préalable – en se laissant toujours une marge de manœuvre – quelles seront les données destinées à fournir une information à ce destinataire. L'enjeu est d'abandonner l'idée que les résultats d'une évaluation puissent servir à informer tous les acteurs, rejoignant là la question de la visée de l'épreuve à décider préalablement à tout développement.

De notre point de vue, nous avons constaté qu'il était possible d'extraire rapidement des informations d'une épreuve informatisée (qu'elle soit plus ou moins adaptative) et que la production automatisée d'un message à l'intention de l'élève et de l'enseignant non seulement était envisageable, mais également souhaitable dans une optique d'épreuve diagnostique nécessitant des remédiations. Cependant, une telle automatisation nécessitera de nombreux ajustements qu'il s'agirait de prévoir, dès les premières conceptions d'EA, par ajustements successifs.

Les nombreuses expériences concernant des épreuves papier, souvent corrigées par les enseignants, nous ont largement renseignés sur les difficultés que présentent une passation équivalente, une correction équitable et une interprétation des résultats cohérente avec les indications fournies. Il serait ainsi préférable de cibler, par épreuve, un nombre restreint de difficultés accompagnées de leurs pistes de remédiation, pariant sur l'évolution des compétences des enseignants à gérer de tels documents au fil des années.

La récupération des résultats, d'une part automatisée pour l'EA, d'autre part auprès des enseignants pour l'EP, est indispensable pour composer l'analyse entre une partie

concernant les ressources et une autre concernant des tâches complexes⁴¹. Ces analyses nécessitent une préparation et un accompagnement scientifique à même de produire différents messages à destination de publics bien ciblés :

- d'abord **pour l'enseignant** afin de lui procurer une aide dans l'interprétation des résultats de l'évaluation. Les enseignants ayant participé à notre expérience, bien qu'intéressés à notre démarche, ont déclaré ne pas pouvoir entrer dans une analyse fine de tableaux de résultats de leurs élèves sans une explication ciblée et claire ; il serait donc préférable de leur fournir prioritairement des messages sur un petit nombre d'élèves (plutôt que sur tous) avec des propositions de prise en charge de leurs difficultés soit par des documents de remédiations individuelles soit par des activités types illustrant des pratiques porteuses des apprentissages attendus ;
- il n'est pas sûr que la synthèse de tous les résultats des classes d'**un établissement** fournisse une véritable aide à sa gestion. Jusqu'à présent, et d'après les pratiques cantonales, les évaluations externes servent avant tout à fournir une information pour le canton et à rassurer les directions d'établissement. Il semblerait approprié de privilégier le repérage de pratiques d'enseignement à améliorer relativement à certains objectifs d'apprentissages du PER, voire de mettre en évidence des ressources trop peu stabilisées chez les élèves ;
- pour **le canton**, les résultats mériteraient de rendre compte du degré de développement des objectifs d'apprentissage dans leur ensemble (et pas seulement d'un cumul de ressources maîtrisées par les élèves) ; ce serait là une façon d'accompagner l'implantation du PER et de construire au fil des années une représentation harmonisée des objectifs.

⁴¹ Notons qu'il serait également possible de tester des tâches complexes sur support informatique.

6. Conclusions

Si, comme nous l'avons vu au chapitre 2 (voir point 2.1), le PER est un curriculum par compétences, cela implique que l'évaluation chargée d'informer du degré d'atteinte de ces compétences porte précisément sur elles. De plus, étant donné la définition même de la compétence (voir point 2.1), il s'agit encore d'évaluer les connaissances et procédures susceptibles d'être mobilisées dans une tâche complexe.

Un modèle d'évaluation en deux phases pour la Suisse romande

Notre modèle d'évaluation convient-il ? Nous avons évoqué plus haut d'autres études menées sur la base des mêmes questions que les nôtres. Au terme de nos lectures, nous avons conclu qu'à l'heure actuelle un modèle d'évaluation comprenant plusieurs phases, chacune mesurant un degré de compétence spécifique (selon les définitions et le modèle de Rey et al., 2004), est le plus abouti pour répondre à nos questions et satisfaire aux contraintes curriculaires du PER.

Que dire de notre modèle construit en deux phases ?

- il mesure des compétences complexes et des connaissances, des procédures (des ressources), celles figurant dans le PER (d'où les deux phases) ;
- il tente d'articuler mesure de connaissances et procédures (ressources) et mesure de compétences, la frontière entre les deux restant perméable et dépendante des élèves ;
- il répond à une visée diagnostique et fournit des résultats interprétables en vue de définir des pistes de remédiation ;
- il convient pour le français comme pour les mathématiques, avec quelques ajustements nécessaires pour répondre aux spécificités de chaque discipline.

Ainsi, nous pouvons affirmer que le modèle d'évaluation développé dans le cadre de ce projet est compatible avec le PER pour les mathématiques et pour le français, et qu'il est en outre adapté à chaque cycle et adaptable à la visée évaluative retenue. Il permet la mesure de connaissances, de procédures et de compétences et en propose une articulation. Les premiers essais de mise en œuvre du modèle montrent la richesse des informations obtenues, ainsi que l'intérêt du modèle lui-même, comme l'attestent par ailleurs les réactions positives de divers partenaires romands à son sujet⁴².

Une fois admis l'intérêt de notre modèle pour l'évaluation des connaissances, procédures et compétences des élèves de Suisse romande, nous devons encore répondre à la question de l'ordre de passation des phases. Nous avons d'abord envisagé de placer en première phase l'évaluation des connaissances et procédures (épreuve adaptative

⁴² Journées d'études des 1^{er} avril 2011, 22 septembre 2011 et 29 mars 2012, à Neuchâtel, organisées par l'IRD; présentations réalisées dans deux centres cantonaux de recherche pédagogique, le SRED à Genève (13 octobre 2011) et l'URSP à Lausanne (9 mai 2012).

en ligne) et en seconde phase l'évaluation de tâches complexes (épreuve papier précédée d'une activité préparatoire). Nous imaginions en effet que les résultats de la phase 1 permettraient la mise en évidence de caractéristiques d'erreurs des élèves en fonction desquelles il aurait été ensuite possible de moduler l'épreuve papier. Dans les faits, et pour des raisons pratiques liées à la vie des classes engagées dans notre investigation, en particulier au matériel informatique disponible, la réalité de la passation a été tout autre. Ainsi certains enseignants ont commencé par l'épreuve adaptative, et d'autres, à l'inverse, par l'épreuve papier. Il est également possible que les deux cas de figure aient eu lieu dans une même classe. Nous avons donc dû considérer les résultats obtenus d'une autre manière, d'abord épreuve par épreuve, puis en tentant de tisser des liens entre elles. Nous pouvons avancer plusieurs arguments privilégiant la réalisation des tâches complexes en premier et celle des tâches portant sur les connaissances et procédures en second :

- cet ordre donne à voir quelles sont les ressources mobilisées ou non par les élèves et l'épreuve adaptative devient ensuite une source d'informations sur les raisons éventuelles d'un usage inadéquat des ressources mobilisées, voire de leur absence;
- Roegiers, lors du colloque de l'ADMEE à Luxembourg au début de 2012, nous a suggéré qu'il n'y avait pas d'arguments scientifiques forts pour décider de l'ordre des phases, mais un argument stratégique : si le curriculum insiste sur les compétences à acquérir, il vaut alors la peine de mettre en avant et en premier la phase qui porte précisément sur les compétences les plus complexes, donc, pour nous, l'épreuve papier.

Pour terminer sur cette question et pour résumer nos constatations, nous affirmons que notre modèle en deux phases, telles que nous les avons construites, est un modèle à partir duquel il est possible d'envisager l'évaluation des compétences des élèves, en référence au PER. Nous estimons encore que ce modèle peut valoir pour toutes les disciplines, étant entendu qu'il permet de prendre en considération leurs spécificités. Nous choisirions aujourd'hui une passation mettant en premier l'évaluation de la tâche complexe, suivie de celle relative aux connaissances et procédures. Enfin, précisons que le type de support n'est peut-être pas aussi exclusif que nous l'avions initialement pensé et que la répartition du support papier pour les tâches complexes et du support informatisé pour les tâches spécifique se discute ; il apparait, plus vivement encore aujourd'hui, que des tâches complexes pourraient tout à fait être réalisées par le biais de l'outil informatique.

Des résultats utiles à l'enseignement-apprentissage

D'autres questions nous paraissent cruciales au terme de notre travail : les résultats obtenus sont-ils utiles ? Quels degrés de compétence peuvent être réellement évalués ? A-t-on vraiment réussi à faire du diagnostic ? Celui-ci constitue-t-il une aide aux enseignants ?

Tout d'abord, nous allons tenter de répondre en nous basant sur les critères développés par De Ketele (2006) et Roegiers (2004) pour évaluer l'instrument de mesure : la pertinence, la validité, la fiabilité et l'efficacité.

La pertinence d'un système d'évaluation renvoie aux fondements qui définissent la réussite. Les tâches d'évaluation éclairent-elles suffisamment les objectifs visés dans le référentiel d'enseignement et permettent-elles de dire quelque chose de pertinent sur le niveau de compétence des élèves évalués ? Ne risque-t-on pas de déclarer des échecs ou des réussites « abusifs » (Roegiers, 2004, p. 164) en ne prenant en considération que quelques objectifs de l'ensemble du référentiel à couvrir ? Autant en mathématiques qu'en français, nous avons veillé à bien définir les éléments du PER que nous voulions évaluer. Les résultats obtenus peuvent très précisément être mis en relation avec ces éléments.

La **validité** du système d'évaluation touche le dispositif d'évaluation lui-même ; celui-ci doit effectivement mesurer ce que l'on veut mesurer. Le choix des questions, des tâches, les critères de correction doivent permettre de dire quelque chose de l'atteinte des objectifs par les élèves et du développement de leur compétence. Le travail en groupes de concepteurs et le fait d'avoir emprunté des questions et des tâches à des épreuves déjà existantes garantissent, à nos yeux, une certaine validité à nos épreuves. Quant aux tâches conçues expressément pour nos épreuves, elles mériteraient une vérification ultérieure de leur validité étant donné que nous n'avons pas pu les pré-tester.

Les conditions de la passation déterminent la **fiabilité** du système d'évaluation. Il faut ainsi éviter que les informations obtenues soient parasitées par des éléments qui ne ressortent pas des performances des élèves. Nous avons été sensibles, par exemple, à la formulation des consignes destinées aussi bien aux élèves qu'aux enseignants. Le travail d'analyse de nos épreuves réalisé par A. Gay (2011) a montré que l'on peut encore améliorer cette formulation et a abouti à des propositions concrètes. Toutefois, sa mise en garde relative à une simplification trop systématique des consignes mériterait d'être approfondie et prise en compte dans les prochaines élaborations de tâches. Notre tendance serait de réserver, pour les apprentissages, des tâches comportant des formulations relativement élaborées afin de familiariser peu à peu les élèves à des consignes toujours plus complexes. Cependant, réduire la complexité des consignes d'une épreuve ne risque-t-il pas, par son effet modélisateur, de décourager le travail sur les consignes réalisés en classe ?

Le principal problème que nous avons rencontré est lié à la passation de l'épreuve papier par les enseignants eux-mêmes. Bien que nous leur ayons fourni un protocole très détaillé de la passation, il est impossible d'être certain que tous ont suivi scrupuleusement les consignes et ne sont pas intervenus auprès de leurs élèves pour répondre à des questions, voire pour donner ici ou là un conseil ou pour faire une observation. Cela tient aussi au fait que les élèves sont très jeunes et ne parviennent pas encore à être totalement autonomes face à la tâche proposée. Nous faisons l'hypothèse que la passation auprès d'élèves des cycles suivants devrait présenter moins de biais de ce type-là. Le mieux serait bien sûr qu'une personne externe, formée, assure la passation de l'épreuve dans la classe. Cette solution est évidemment coûteuse.

La question de l'**efficacité** du système d'évaluation est particulièrement délicate. La perspective diagnostique qui a été la nôtre produit-elle des résultats utiles pour les enseignants ? Disposons-nous d'informations précises et dont le sens est suffisamment clair pour que les enseignants puissent organiser leur travail à partir d'elles ? Nous avons vu, dans l'analyse de nos résultats, que les épreuves – aussi bien en français qu'en mathématiques – fournissent des informations très nombreuses et fines sur

les compétences des élèves. Plus encore, elles mettent en avant les principales difficultés que ceux-ci rencontrent au fil des questions et des tâches. Reste à savoir dans quelle mesure les difficultés recensées constituent des obstacles à la suite des apprentissages, et doivent donc être prises en considération. Nous avons écarté par exemple les difficultés liées à des attentes fondamentales, trop éloignées encore des possibilités d'élèves de début de 5^e. Du fait que nous avons dû confectionner des épreuves adaptées à l'âge des élèves, donc courtes, nous disposons de trop peu d'informations pour être sûrs que les difficultés rencontrées par certains élèves auraient été les mêmes dans des tâches similaires à celles de nos épreuves. Autrement dit, avons-nous assez de questions et de tâches pour affirmer que telle difficulté reflète l'état de connaissance ou de compétence d'un élève ?

La question des remédiations à proposer pour tel ou tel type de difficulté se pose encore. Ces remédiations vont-elles véritablement permettre aux élèves de surmonter leurs difficultés ? Les concepteurs des épreuves n'ont-ils pas tendance à croire un peu trop facilement qu'à telle difficulté correspond telle remédiation, celle-ci effaçant alors celle-là ? Ne devrait-on pas laisser au système sa souplesse qui permet à l'enseignant d'agir comme il le veut, quitte à ne pas recourir aux pistes de remédiation proposées et à réguler son enseignement d'une autre manière ? Nous avons vu qu'en mathématiques, la remédiation peut accompagner des difficultés très fines liées à des questions ou des tâches ciblées en fonction d'objectifs ou d'attentes fondamentales précises. Cette approche micro peut présenter un risque d'enfermement de l'apprentissage dans un pas à pas contraignant, contraire finalement à l'esprit même de la discipline. Il s'agirait de mieux repérer ce qui doit être nécessairement acquis et ce qui peut encore se déployer dans le temps, selon des démarches souples et multiples. En revanche, en français, le problème est inverse. L'approche macro qui a été développée pour cette discipline présente le risque de faire apparaître des types généraux de difficultés, sans surprise pour des enseignants, que des remédiations ponctuelles ne suffiraient pas à aplanir. Nous avons vu également que les pistes fines et précises de remédiation, en particulier pour la compréhension de textes, ne sont pas légion.

Nous avons réalisé une évaluation à visée diagnostique externe, conçue pour tous les élèves des classes concernées, afin de nous permettre avant tout de valider notre modèle, mais aussi de repérer des types de difficultés. Or, dans la réalité des classes, il serait certainement plus utile d'envisager une évaluation diagnostique souple, conduisant les enseignants à ne faire passer les épreuves (ou certaines parties d'entre elles) qu'aux élèves pour lesquels ils voudraient obtenir des informations plus fines que celles que leur procure leur enseignement. Il n'empêche, qu'à l'instar de Rey et al. (2004), nous pouvons admettre que toute évaluation à une portée diagnostique dès lors qu'elle fournit aux enseignants des informations utiles pour leur enseignement. Il n'est donc pas exclu de l'envisager à large échelle pour tous les élèves. De plus, son effet modélisateur pourrait amener peu à peu les enseignants à changer leurs pratiques d'évaluation en classe et jouer ainsi un rôle formateur.

Nous voulons encore aborder l'aspect **réaliste** du système d'évaluation que propose Roegiers (2004). Ce point touche à la facilité d'élaboration, de correction, voire de notation des épreuves. Pour ce qui est de l'élaboration des épreuves (adaptative, en ligne et papier), il faut bien admettre qu'elle n'est pas aisée. Choisir une tâche, des questions, des distracteurs, élaborer des consignes, posent à chaque fois des questions cruciales auxquelles seul un groupe de personnes rodées (spécialistes de l'évaluation, de la statistique, didacticiens, enseignants, formateurs, etc.) peut répondre. Même

pour ces spécialistes, l'élaboration de tâches complexes d'évaluation n'est guère aisée. La distinction entre ces tâches et des tâches complexes destinées à l'apprentissage est à assurer. Globalement, une tâche complexe destinée à l'enseignement doit offrir la possibilité à l'élève de réinterroger ses connaissances et sa façon de gérer l'organisation d'une résolution ou d'une production, alors qu'une tâche complexe d'évaluation n'a pas pour objectif de déstabiliser ou de désorganiser les ressources des élèves mais bien de les leur faire mobiliser et combiner. Bien entendu, la nuance est ténue, sachant qu'il est nécessaire d'estimer les ressources que l'on peut raisonnablement considérer comme acquises pour une population d'élèves.

Des biais sont si vite introduits dans l'épreuve que sa construction exige une très forte attention et beaucoup de vigilance, la correction n'étant guère plus aisée. Les guides de correction doivent être rédigés de manière la plus claire possible. Évidemment, la correction d'une tâche complexe, par exemple d'un texte produit par les élèves, est de ce point de vue difficile à mettre en place. La part d'interprétation des correcteurs est plus grande qu'imaginée au départ, même si le libellé proposé est le plus précis possible. Faut-il alors recourir à plusieurs correcteurs ? Dans une perspective diagnostique, ce n'est pas ce qui est souhaité mais plutôt que l'enseignant lui-même puisse assurer cette correction puisque c'est lui qui devrait en tirer bénéfice. Toutefois, une longue correction peut s'avérer délicate et dissuasive pour ce dernier. Le caractère réaliste du système d'évaluation est finalement l'un des plus difficiles à assurer !

En citant encore la **visibilité** de ce système, Roegiers (2004) souligne l'importance de l'information qu'il devrait permettre de donner à toutes les instances concernées. Comment pouvons-nous informer les enseignants de nos résultats et de leurs conséquences pour leur enseignement ? Nous avons parlé plus haut des pistes de remédiation, tout en nuancant leur portée, en particulier en français. Mais au-delà des pistes de remédiation, nous savons que d'autres documents ou d'autres types d'informations seraient également à développer : auprès des formateurs par exemple, pour les aider à accompagner les enseignants dans le processus d'évaluation du travail de leurs élèves en classe ; auprès des enseignants aussi, pour qu'ils comprennent ce qui est visé par l'évaluation et comment ils peuvent eux-mêmes en faire un outil utile pour leur enseignement. Nous n'avons pas pu explorer tous les canaux d'information aux personnes susceptibles d'être intéressées par les résultats d'une évaluation diagnostique. Nous pensons par exemple aux parents qui suivent souvent les apprentissages de leur(s) enfant(s). Ce travail reste à faire.

Les avantages d'une évaluation diagnostique

Rappelons que le choix d'une visée diagnostique pour nos épreuves prototypes présentait plusieurs avantages dans le cadre de notre recherche :

- La démarche étant exploratoire, il n'était pas question de prétendre aboutir à une évaluation sommative finalisée et en parfaite adéquation avec le PER, qui plus est au moment où ce dernier se met en place.
- L'intérêt d'un diagnostic nous semblait plus porteur pour les enseignants qui souvent, se sentent peu investis dans des évaluations systèmes ; celles-ci risquent en effet de démobiliser élèves et enseignants, ce que nous souhaitons éviter consi-

dérant qu'ils devraient être les partenaires privilégiés dans l'élaboration de telles épreuves⁴³.

- L'évaluation de compétences est un domaine de recherche extrêmement actuel et, par conséquent, en mouvement; elle s'étend de l'évaluation individuelle à l'évaluation système (comme PISA), les caractéristiques de l'une n'étant pas transférables telles qu'à l'autre. Il nous paraissait important de porter quelque lumière sur l'individuelle, encore largement à stabiliser.
- Une épreuve diagnostique permet de développer des tâches fines et adaptées aux objectifs du PER (cf. chapitre 3.3.3). Elle permet encore d'analyser en profondeur les résultats obtenus dans la mesure où les tâches ne sont pas obligatoirement nombreuses pour une épreuve, le but étant avant tout de repérer les principales difficultés des élèves et non de couvrir l'ensemble des objectifs de la fin d'un cycle. Nous avons également montré que ces tâches pouvaient ensuite être versées dans une banque d'items, et que les éléments de leur classification apportaient des informations précises et indispensables aux concepteurs d'épreuves. Une telle banque, nous l'avons dit, peut rassembler des tâches provenant de tous les cantons et rapidement contenir un matériel disponible. Sur cette base substantielle, il deviendrait alors possible de construire des épreuves à visées autres que diagnostique.

⁴³ Déjà en 1954, l'équipe de Bloom (1969) constatait : (...) *une des plus sérieuses difficultés relatives à la classification des questions d'examen : la nécessité de connaître ou de présumer la nature des expériences pédagogiques antérieures de chaque sujet examiné. (...) En conclusion, le matériel d'examen ne peut être classé de façon satisfaisante que s'il est possible de connaître ou de présumer les conditions dans lesquelles se présentent les candidats.* (p. 24)

7. Les perspectives

La mutualisation des expériences cantonales en matière d'évaluation est nécessaire pour l'élaboration de questions ajustées, pour la validation du modèle d'évaluation et pour une interprétation homogène des éléments du PER. Elle présente le double avantage d'accompagner la mise en place de ce référentiel dans les cantons et d'alimenter une banque d'items régionale. Dans cette optique, il nous paraît utile d'établir une sorte de fil rouge de ce qui pourrait constituer la suite d'un travail scientifique d'élaboration d'épreuves pour la Suisse romande.

Il apparaît clairement que les épreuves d'évaluation externe jouent souvent un rôle modélisant auprès des enseignants. Ce rôle rend le travail de réalisation des épreuves encore plus exigeant et met leurs concepteurs devant d'importantes responsabilités. Il s'agit alors de mettre en évidence les enjeux majeurs des apprentissages scolaires propres à chaque discipline, voire transversaux. Dans ce sens, les attentes fondamentales fournissent des repères utiles. Cependant, nous l'avons vu, les objectifs d'apprentissage et leurs composantes ouvrent l'espace riche mais difficile à couvrir qu'est celui des tâches complexes. Des critères d'analyse clairs, pertinents, opérationnels de ces tâches sont nécessaires pour la compréhension de ce qui est visé par l'évaluation. L'évaluation représente donc également un enjeu majeur de la mise en œuvre du PER. Le modèle d'évaluation arrêté doit pouvoir garantir un juste équilibre entre les diverses dimensions du PER. C'est pourquoi, si les perspectives présentées ci-après constituent un outil de travail qui nous semble indispensable, cet outil doit absolument être discuté et faire l'objet d'une réappropriation par tous les acteurs engagés dans l'élaboration d'épreuves pour la Suisse romande. Rappelons que ces perspectives s'inscrivent dans le cadre du modèle d'évaluation en deux phases que nous avons développé tout au long de notre travail.

7.1 Les conditions d'élaboration d'une évaluation

Avant d'entreprendre la construction d'une évaluation, il est nécessaire de savoir quelle en est la visée. En effet, sa structure n'est pas la même s'il s'agit d'orienter les élèves (visée pronostique) ou plutôt d'en évaluer les acquisitions au terme d'un cycle (visée sommative), si l'on veut poser un diagnostic en début de cycle (visée diagnostique) ou encore si c'est l'efficacité du système scolaire qui est mesurée (visée de bilan).

Dans le cas de l'orientation, à côté de tâches visant à connaître les acquis indispensables des élèves, on sera tenté de proposer des tâches de difficultés diverses et touchant aux attentes du cycle suivant. La visée sommative d'une épreuve conduira, quant à elle, à estimer dans quelle mesure les élèves ont bien atteint les objectifs attendus en fin de cycle.

Pour ce qui est du diagnostic, il peut s'adresser à certains élèves seulement, et l'épreuve devrait servir à mettre en avant les principales difficultés rencontrées par les élèves,

susceptibles de les empêcher d'avancer dans leurs apprentissages, nécessitant donc des remédiations préalables.

Enfin, le système peut se contenter de faire passer l'épreuve à des échantillons d'élèves afin de savoir dans quelle mesure la majorité des attentes ou des compétences fondamentales sont acquises par la population concernée. Les tâches seront donc nombreuses mais réparties dans des cahiers, eux-mêmes répartis sur la population testée; ainsi, toutes les tâches ne seront pas présentées à tous les élèves car ce ne sont pas les résultats individuels qui sont alors recherchés mais le taux de réussite aux objectifs.

Pour le cas où tous les élèves devraient être évalués (comme le préconise la CSR), il faudrait prendre certaines contraintes en considération :

- Le nombre de tâches ne peut être trop important (surtout pour de jeunes élèves), ce qui limite le champ des attentes ou des objectifs que l'on veut couvrir. Il s'agira donc de choisir soigneusement ces tâches, afin que les résultats obtenus puissent tout de même renseigner sur les acquisitions fondamentales attendues.
- La passation devrait se faire pour tous en même temps (ou presque) et selon des conditions qui n'introduisent pas de biais dans les résultats; dans le cas où les enseignants assurent la passation, il est important de leur fournir des consignes d'une grande précision pour éviter, dans la mesure du possible, des interventions intempestives de leur part qui pourraient modifier les résultats.
- La question de la correction des épreuves est également délicate; là encore, si ce sont les enseignants qui l'assurent, elle doit être aisée, claire pour chacun, ne pas prendre trop de temps et ne pas donner lieu à des interprétations risquant d'être différentes d'un évaluateur à l'autre. Le matériel utilisé doit être adapté aux possibilités de toutes les écoles de tous les cantons.

1^{re} proposition : déterminer les spécificités d'une évaluation en fonction de sa visée et des contraintes (organisationnelles, temporelles, ...) dont il faut tenir compte sans que cela oblitère l'évaluation de compétences.

7.1.1 Pistes pour l'élaboration d'épreuves

Avant de définir les caractéristiques des tâches (spécifiques, complexes, complémentaires), il est nécessaire de connaître l'ampleur des évaluations attendues, de définir les objets de la mesure, ainsi que l'information qui doit en être extraite. Ces clarifications apportées, les questions didactiques disciplinaires, en lien avec le référentiel d'enseignement, doivent trouver des réponses quant aux éléments qu'il est effectivement possible de mesurer, à la variété des tâches portant sur un même élément, à l'équilibre entre ces tâches au sein de chaque partie de l'évaluation (que ce soit deux épreuves ou plus), à l'interprétation des difficultés observées dans la réalisation de chaque épreuve et aux pistes de remédiation qu'elle ouvre. La question de l'ampleur d'une tâche complexe demeure intacte : une tâche complexe peut-elle s'identifier à une épreuve ?

Se pose ainsi la question des liens entre les deux phases de l'évaluation. Il est important de comprendre ce qui fonde ces liens. L'ordre de passation de ces phases reste ouvert : ces différentes questions méritent non seulement d'être discutées mais également de donner lieu à des explorations dans les classes.

2^e proposition : clarifier, selon les orientations données pour les évaluations, quelles sont les caractéristiques des parties constitutives de l'évaluation et les liens qu'elles tissent entre elles.

7.1.2 Conditions à assurer pour une épreuve informatisée

La conception des épreuves adaptatives (pour les mathématiques et pour le français) a constitué pour nous une expérience nouvelle et riche. En les centrant sur la vérification de connaissances et de procédures susceptibles de tenir lieu de ressources pour des tâches complexes, nous avons construit des questions que l'on trouve habituellement dans les tests de connaissance, dans les épreuves généralement réalisées au sein de l'école.

La question se rapporte à un objectif d'apprentissage, à une ou plusieurs attente(s) fondamentale(s), et se décline en une consigne et une tâche à réaliser. Il nous a toutefois semblé intéressant d'essayer de tirer parti de l'outil informatique pour obtenir des informations que seul un tel outil peut fournir. Nous avons imaginé ainsi des tâches permettant des manipulations spécifiques (placer en glissant des objets, placer des séparateurs entre des mots, cliquer, etc.). Il devient alors possible, si on l'estime important pour comprendre comment a procédé l'élève, de « photographier » (garder une trace graphique) les manipulations effectuées par ce dernier, ce qui, dans une perspective diagnostique, peut s'avérer particulièrement intéressant.

Un autre apport de l'outil informatique est la mesure exacte du temps passé pour répondre à une question. Grâce à cette mesure, on a pu par exemple mettre en évidence en mathématiques une attitude de zapping, montrant que certains élèves traversent les questions en s'y arrêtant trop peu de temps pour y répondre correctement. En français, nous avons mesuré le temps pris par chaque élève pour lire le texte au départ des questions de compréhension. De plus, nous avons pu repérer, pour plusieurs questions, les retours des élèves au texte de départ et pour combien de temps. Temps de lecture du texte et retours au texte ont pu ensuite être associés aux résultats obtenus. Toujours en français, nous avons donné la possibilité aux élèves de pouvoir écouter la consigne des questions pour le cas où la lecture de consignes risquait de représenter une charge cognitive trop lourde pour ceux éprouvant des difficultés en lecture.

Ces exemples montrent qu'il est possible, grâce à l'outil informatique, de recueillir des informations utiles venant compléter les résultats du test lui-même. Ces informations peuvent fournir des renseignements touchant aux procédures et stratégies de travail des élèves que nous n'aurions pas obtenus par le biais d'une épreuve traditionnelle sur papier.

L'outil informatique permet en outre d'envisager une épreuve adaptative, c'est-à-dire tenant compte de certaines réponses à des questions spécifiques pour conduire les élèves vers une question plus facile ou plus complexe. Ce parcours individualisé devrait, par exemple dans une perspective diagnostique, offrir la possibilité d'établir des profils de réponses types à même d'aider les enseignants à proposer à leurs élèves des activités de remédiation en fonction de ces profils. Dans une évaluation sommative, ces parcours types permettent de mieux comprendre et d'affiner le degré d'acquisition des connaissances par les élèves.

Nous avons opté pour une épreuve adaptative en ligne mesurant les compétences de premier ou deuxième degré selon Rey et al. Mais il nous semble tout à fait possible d'utiliser l'outil informatique pour évaluer des tâches complexes correspondant à la mesure de compétences du troisième degré. Prenons le cas de la lecture. Il serait tout à fait possible, en particulier grâce à la navigation sur Internet, de faire chercher toutes sortes d'informations en lien avec un texte de départ (à lire ou à écrire), par exemple, et en fonction d'un but particulier assigné à cette lecture (faire un exposé, organiser une sortie, enrichir une lecture littéraire en cherchant à mieux connaître l'écrivain, le contexte historique, etc., compléter les informations d'un article scientifique, ...). Dans le cadre de l'étude PISA (2011), une épreuve en ligne a été effectuée en 2009 (à laquelle la Suisse n'a pas participé) montrant effectivement tout l'intérêt de l'informatique pour obtenir des informations riches sur la compétence en littérature des jeunes de 15 ans.

3^e proposition : développer la création d'épreuves informatisées pour évaluer les compétences de tout degré et en définir leurs caractéristiques optimales.

7.2 Les conditions d'élaboration de tâches

7.2.1 Pistes pour la rédaction de tâches

Si le modèle d'évaluation distingue nettement ce qui relève de tâches spécifiques (les ressources à mobiliser dans des tâches complexes, des connaissances ponctuelles) et de tâches complexes, dès qu'il s'agit de construire une épreuve selon ces deux catégories de tâches, la distinction n'est pas toujours aussi simple (cela est par exemple apparu pour la compréhension en lecture : qu'est-ce qui relève explicitement des ressources à mobiliser ? De la connaissance du code linguistique ? Des connaissances grammaticales ? Des stratégies de lecture ? A quel moment, lorsqu'on aborde ces questions, quitte-t-on les tâches spécifiques pour entrer dans une tâche complexe ?). Comment, dans une tâche complexe, s'assurer que telle ou telle ressource a bien été utilisée ? Et à quel degré de compétence se réfère-t-elle ? Un premier constat est que les degrés de compétence utilisés dans notre recherche présentent une piste intéressante pour faire de l'ordre dans les éléments que l'on souhaite mesurer, mais nécessitent des discussions en présence de chercheurs et d'acteurs du terrain.

Les tâches complexes méritent aussi des clarifications quant à leur ampleur (voir point 7.1.1), à l'accessibilité de la situation présentée aux élèves, aux ressources à mobiliser, etc. Déterminer un temps de passation pourrait être une piste intéressante pour contenir leur empan, et pour permettre une prise en compte des temps de concentration des élèves.

Dans la perspective diagnostique qui a été la nôtre, l'élaboration de tâches complémentaires pour l'EA a été envisagée. Elle s'avère délicate car il s'agit, pour chaque tâche complémentaire, de s'assurer :

- qu'elle présente un niveau de difficulté plus bas que la précédente (ou de celle pour laquelle on la propose) ;
- qu'elle permette bien d'identifier des obstacles ou difficultés ayant pu produire un échec à la tâche à laquelle elle succède ;

- qu'elle n'avantage pas les élèves qui la passent pour le cas où elle présenterait une situation similaire à celle d'une tâche obligatoire.

Il est certain qu'une analyse statistique tel que l'IRT constitue une aide pour résoudre cette question de l'ordre de difficultés de plusieurs tâches. Toutefois, il ne s'agirait pas d'exclure des tâches trop simples censées précisément repérer les exceptionnels cas d'élèves en difficulté sur certains objets.

Les tâches complémentaires prennent sens dans une évaluation diagnostique et également dans une évaluation pronostique (en particulier pour orienter les élèves dans des filières ou des niveaux; elles peuvent par exemple distinguer ceux qui réussissent des tâches plus difficiles que celles attendues et donc destinés à des filières ou niveaux à exigences élevées).

Le rôle d'une grille d'analyse de consignes, comme A. Guay (2011) le proposait dans son rapport, mérite réflexion : doit-elle intervenir en amont ou en aval des rédactions ?

4^e proposition : s'accorder sur les caractéristiques des tâches complexes, des tâches spécifiques et complémentaires; si possible, définir en consortiums romands les ressources mobilisées (et issues du PER) à travers ces tâches.

7.2.2 *Éléments de classification*

Dans une optique de mutualisation des pratiques évaluatives cantonales et romandes, l'IRD⁴⁴ s'est donné pour tâche la conceptualisation et la réalisation d'une banque d'items régionale. Au départ, elle a regroupé les épreuves sous forme papier, à la disposition des cantons. Dans le cadre du mandat EpRoCom, le projet Prosper (Projet Outils de Suivi du PER) poursuit cette tâche. Son objectif est, dans un premier temps, de créer un fonds commun, accessible à tous les cantons, des épreuves et des tâches élaborées en partenariat avec ces derniers; des épreuves cantonales pourront y être insérées selon des modalités discutées au niveau romand. Il n'est pas exclu, dans un deuxième temps, d'ouvrir une partie de la banque aux enseignants. Bien évidemment, tout ceci reste encore à définir par les autorités compétentes.

Le processus de mutualisation à mener et à définir avec les cantons implique une classification bien précise des épreuves et des tâches sur laquelle nous avons dû nous pencher dans le cadre des présents travaux (voir annexes points 8.3 et 8.4). Il n'est toutefois pas opportun, dans ce rapport, de fournir une explication de chacun des critères de cette classification. Seules une mise en évidence et une justification de ceux retenus seront discutées. Ces critères sont issus du modèle de classification élaboré par Pochon (2007) et des exemples concrets permettent de saisir quels sont les éléments essentiels de classification des tâches et des épreuves. Soulignons encore l'intérêt du modèle du projet EVAPM⁴⁴ dont le but est de créer une base de données statistiques ainsi que la mise à disposition de questions d'évaluation. Toutes les expériences relatées par Pochon (2006, 2007) ont permis de définir des critères globaux à prendre en compte pour classer des problèmes⁴⁵ et soulève la question subsidiaire

⁴⁴ Évaluation des apprentissages mathématiques.

⁴⁵ Le terme « problèmes » correspond à « tâche ».

– et pas des moindres – des ontologies⁴⁶. Trois entrées sont proposées : description de l'activité (énoncé, consigne, forme des observables), classification selon le domaine (paramètres, niveaux, analyse, compétences) et documentation (historique, auteur, analyse d'erreurs, indices statistiques).

Par souci d'alléger la lecture, nous avons fait le choix de retenir dix éléments de classification des tâches, dont deux (signalés par une *) concernent uniquement l'épreuve adaptative :

- **références au PER** : se réfèrent aux objectifs d'apprentissage et aux attentes fondamentales associées ;
- **fonctionnalités*** : concernent les fonctionnalités informatiques des tâches et doivent être clarifiées avant tout développement ;
- **variables et réponses construites** : correspondent aux données prélevées, certaines automatiquement dans l'EA, puis reconstruites à la suite des observations effectuées ;
- **ce que l'on mesure** : se distingue du référentiel PER par l'aspect évaluatif plus fin qu'il apporte (p. ex. pour M03 : *repère les nombres ayant 3 comme chiffre des unités* fait appel aux attentes fondamentales suivantes : *compte de 10 en 10 jusqu'à 100 à partir d'un nombre donné et compare, (...), intercale des nombres inférieurs à 10*) ;
- **obstacles a priori** : rendent compte d'une mauvaise démarche possible de l'élève, notamment en fonction de certains éléments de la tâche ;
- **difficultés observées et/ou possibles** : détaillent les difficultés estimées avant (en lien avec les obstacles) ou après la production (en fonction alors des observations effectuées) ;
- **résultat attendu** : décrit la (les) réponse(s) et/ou valeur(s) correspondant à la réussite de la tâche ;
- **aiguillage*** : indique une navigation dans l'EA, avec ou sans embranchement ;
- **variables didactiques** : rendent compte des possibilités de faire varier la tâche en jouant sur les paramètres ;
- **consigne pour l'enseignant** : réservée *a priori* aux tâches écrites, elle indique l'encadrement de la tâche par l'enseignant, éventuellement sa marge de réponse face à une question (partie encore à développer) ;
- **améliorations** : portent sur la tâche, autant dans sa forme, sa consigne, ses paramètres, etc., en fonction de certaines difficultés ou de mauvaises manipulations repérées.

Les éléments de classification retenus pour décrire les épreuves sont les suivants :

- **références au PER** : concernent les objectifs d'apprentissage et les attentes fondamentales associées, avec un cumul du nombre de tâches concernées ;

⁴⁶ Étude des termes constituant les rubriques de classification et des catégories possibles pour chaque rubrique.

- **visée de l'épreuve**: en lien avec les définitions données, elle peut être soit diagnostique, sommative, d'orientation, bilan, etc.; dans notre cas, il s'agit d'une visée diagnostique;
- **description de l'épreuve**: détaille les tâches constitutives de l'épreuve, l'ordre, etc.;
- **déroulement de l'épreuve**: détaille la façon dont l'épreuve se déroulera en classe;
- **variables et réponses contruites sur l'ensemble de l'épreuve**: il peut s'agir soit de variables comprenant des informations sur l'élève, soit de variables construites sur l'ensemble de l'épreuve par la prise en compte de plusieurs variables (comme par exemple les critères décrits aux points 5.2.2, 5.2.3 et 5.2.4).

En lien avec les travaux préalables de Pochon (2006; 2007), nous constatons que nos choix des éléments de classification sont cohérents avec ceux qui y sont explicités. Il s'agit en effet de décrire la tâche ainsi que l'épreuve dans laquelle elle s'inscrit, de mentionner les compétences mobilisées (référence au PER), de préciser un certain nombre de paramètres (variables didactiques), et d'analyser les difficultés observées et/ou possibles, les obstacles *a priori* et les réponses.

Le modèle de classification que nous avons élaboré a été mis à l'épreuve non seulement lors de la classification des épreuves et des tâches EpRoCom mais également lors de la classification d'épreuves de mathématiques du cycle 3 réalisées par V. Marc et B. Vermot⁴⁷. Par ailleurs, lors de la journée d'étude du 29 mars 2012 organisée par l'IRD, *Caractériser une classification commune des tâches et des épreuves*, il nous a été possible de confronter nos éléments de classification aux réalités cantonales et de les ajuster. Cette journée a servi à clarifier quelles seraient les informations incontournables à recueillir afin de permettre une mise en commun des pratiques d'évaluations cantonales. Plus précisément, il ne s'agissait pas de traiter du contenu des épreuves et des tâches mais des rubriques sous-jacentes à ces dernières. Si la majorité des éléments de classification considérés comme incontournables faisait consensus, celui concernant « ce que l'on mesure » a suscité quelques débats: certains ont en effet proposé de le supprimer, considérant que la rubrique *référence au PER* devait suffire. D'autres au contraire l'ont jugé indispensable pour les concepteurs d'évaluations, insistant sur l'affinement absolument nécessaire du référentiel en termes évaluatifs.

5^e proposition: définir les éléments de classification des tâches permettant de répertorier, d'exploiter et d'améliorer les tâches communes, et caractérisant la mutualisation possible des tâches et épreuves cantonales.

7.2.3 Vers une banque d'items régionale

Notre travail ainsi que les journées d'études sur l'évaluation en référence au PER ont montré l'importance d'une collaboration intercantonale pour élaborer des épreuves romandes; la création d'une banque d'items pourrait alors offrir un outil de travail et un réservoir de tâches et d'épreuves dans lequel puiser pour répondre aux besoins d'évaluation régionale.

⁴⁷ Dans le cadre d'un mémoire de Bachelor de Psychologie de l'Université de Fribourg.

Qu'entendre par banque d'items? A l'IRD, nous proposons la définition suivante⁴⁸: *On désigne par ce terme un répertoire (liste, catalogue, recueil) d'items qui, généralement, ont déjà été utilisés à des fins d'évaluation et qui sont mis à la disposition de personnes (chercheurs ou enseignants) souhaitant y avoir recours (accès libre ou soumis à certaines conditions)*. Selon Chartier et al. (2000) on peut, en éducation, parler aussi de banque d'instruments, de banque d'outils d'évaluation voire de banque d'exercices. Cette banque peut répondre à des attentes multiples, à des visées évaluatives différentes. Par ailleurs, dans notre cas, il ne faut pas comprendre « items » dans un sens restreint de « questions » ; il peut s'agir également de « tâches », qu'elles soient spécifiques ou complexes.

Si une telle banque devait se développer en Suisse romande, il conviendrait de définir les conditions de son exploitation : que veut-on y mettre ? Qui en aurait l'accès ? Quel suivi est-il prévu pour qu'elle soit à jour ? Quel lien pourrait-elle avoir avec la banque nationale prévue dans le cadre d'HarmoS ? Il est important, si l'on veut qu'un tel instrument soit au service de la Suisse romande, de répondre avec soin à de telles questions et d'assurer un suivi par une instance spécialisée et ayant les forces de le réaliser.

6^e proposition : développer une banque d'outils d'évaluation romande en tenant compte des spécificités du PER et du modèle d'évaluation proposé dans ce rapport.

⁴⁸ <http://www.irdp.ch/edumetrie/lexique/item.htm>

8. Annexes

8.1 Liste des membres participants

Membres du Consortium romand

BE Riesen Werner
FR Gremaud Nicole
GE Rouget Philippe
GE Ntamakiliro Ladislas
JU Bréchet Michel
NE Kleiner Claude-Alain
VS Métrailler Denis
VD Guillod Philippe

Membres des groupes de travail

En français :

Beaud Jean-François
Broi Anne-Marie
Merkelbach Christian
Pfeiffer-Ryter Véronique
Wacker Muriel
Wirthner Martine

En mathématiques :

Calame Jacques-André
Dumas Jean-Paul
Guignard Ninon
Marc Viridiana
Pochon Luc-Olivier
Riesen Werner
Vermot Bertrand

8.2 Description des évaluations cantonales par cycle

Avertissement : la dénomination « épreuve cantonale » est utilisée dans les tableaux ci-dessous autant pour parler d'épreuves de référence que d'épreuves communes ou cantonales ; en effet, l'intention est avant tout de mettre en évidence quelles sont les épreuves constituées par le canton et celles qui ne le sont pas (p. ex. celles constituées par un établissement).

Évaluations cantonales au cycle 1 (état 2012)

	Cycle 1			
	1 EE	2 EE	3 ^e (ex-1P)	4 ^e (ex-2P)
BE				
FR				épreuve cantonale en : - français - ou mathématiques (en alternance avec 6 ^e) <small>0 0 10 11 12 1 2 3 4 5</small>
GE				épreuves cantonales en : - français - écriture-graphisme - mathématiques <small>0 0 10 11 12 1 2 3 4 5</small>
JU				
NE			épreuve cantonale dans une discipline (français ou math. ou musique ou...); (discipline identique pour les 3 ^e à 7 ^e années) <small>0 0 10 11 12 1 2 3 4 5</small>	épreuve cantonale dans une discipline (français ou math. ou musique ou...); (discipline identique pour les 3 ^e à 7 ^e années) <small>0 0 10 11 12 1 2 3 4 5</small>
VS				épreuve cantonale en : - français - mathématiques - environnement <small>0 0 10 11 12 1 2 3 4 5</small>
VD				épreuves cantonales en : - français - mathématiques <small>0 0 10 11 12 1 2 3 4 5</small>

Épreuve à visée sommative : prise en compte dans la promotion de l'élève
 Épreuve à visée diagnostique : résultats indicatifs
 Épreuve inscrite dans la loi scolaire du canton

en jaune : mois de passation de l'épreuve
0 0 10 11 12 1 2 3 4 5

Évaluations cantonales au cycle 2 (état 2012)

	Cycle 2			
	5 ^e (ex-3P)	6 ^e (ex-4P)	7 ^e (ex-5P)	8 ^e (ex-6P)
BE				épreuves par zone de recrutement sec. I en : - français - allemand - mathématiques <small>8 9 10 11 12 1 2 3 4 5 6</small>
FR		épreuves cantonales en : - français - mathématiques (en alternance avec 4e) <small>8 9 10 11 12 1 2 3 4 5 6</small>		épreuves cantonales (PPO) en : - français - mathématiques - allemand - connaissances gén. <small>8 9 10 11 12 1 2 3 4 5 6</small>
GE		épreuves cantonales en : - français I (communic.) - français II (structuration) - mathématiques - allemand <small>8 9 10 11 12 1 2 3 4 5 6</small>		épreuves cantonales en : - français I (communic.) - français II (structuration) - mathématiques - allemand <small>8 9 10 11 12 1 2 3 4 5 6</small>
JU		épreuves cantonales en : - français - mathématiques (tous les deux ans) <small>8 9 10 11 12 1 2 3 4 5 6</small>		épreuves cantonales en : - français - mathématiques - allemand <small>8 9 10 11 12 1 2 3 4 5 6</small>
NE	épreuve cantonale dans une discipline (français ou math, mus, ...); (disc. identique pour les 3 ^e à 7 ^e années) <small>11 12 1 2 3 4 5 6</small>	épreuve cantonale dans une discipline (français ou math, ou musique ou...); (discipline identique pour les 3 ^e à 7 ^e années) <small>8 9 10 11 12 1 2 3 4 5 6</small>	épreuve cantonale dans une discipline (français ou math, mus, ...); (disc. identique pour les 3 ^e à 7 ^e années) <small>11 12 1 2 3 4 5 6</small>	épreuves cantonales en : - français - mathématiques - allemand (2 ^e session) <small>8 9 10 11 12 1 2 3 4 5 6</small>
VS		épreuves cantonales en : - français - mathématiques <small>8 9 10 11 12 1 2 3 4 5 6</small>		épreuves cantonales en : - français - mathématiques <small>8 9 10 11 12 1 2 3 4 5 6</small>
	épreuve cantonale en : - allemand - environnem. <small>11 12 1 2 3 4 5 6</small>	épreuve cantonale en : - allemand <small>8 9 10 11 12 1 2 3 4 5 6</small>	épreuve cantonale en : - allemand - sciences <small>11 12 1 2 3 4 5 6</small>	épreuve cantonale en : - allemand <small>8 9 10 11 12 1 2 3 4 5 6</small>
VD		épreuves cantonales en : - français - mathématiques <small>8 9 10 11 12 1 2 3 4 5 6</small>		épreuves cantonales en : - français - mathématiques - allemand <small>8 9 10 11 12 1 2 3 4 5 6</small>

Épreuve à visée sommative : prise en compte dans la promotion de l'élève
Épreuve à visée diagnostique : résultats indicatifs
Épreuve inscrite dans la loi scolaire du canton

en jaune : mois de passaton de l'épreuve
8 9 10 11 12 1 2 3 4 5 6

Évaluations cantonales au cycle 3 (état 2012)

	Cycle 3		
	9 ^e (ex-7 ^e)	10 ^e (ex-8 ^e)	11 ^e (ex-9 ^e)
BE			
FR	épreuves cantonales en : - français - mathématiques - allemand - anglais	épreuves cantonales en : - français - mathématiques - allemand - anglais	épreuve cantonale (Diplôme de fin de scolarité oblig.) en : - français - mathématiques - allemand - anglais (- latin)
	8 9 10 11 12 1 2 3 4 5	8 9 10 11 12 1 2 3 4 5	8 9 10 11 12 1 2 3 4 5
GE	épreuves cantonales en : - français - mathématiques - allemand - latin (pour les MA-latin)	épreuves cantonales en : - français - mathématiques - allemand - anglais - latin - physique - biologie	épreuves cantonales en : - français - mathématiques - allemand - anglais - latin - physique
	8 9 10 11 12 1 2 3 4 5	8 9 10 11 12 1 2 3 4 5 6	8 9 10 11 12 1 2 3 4 5 6
JU		épreuves cantonales en : - français - mathématiques (tous les deux ans)	
		8 9 10 11 12 1 2 3 4 5 6	
NE		épreuves cantonales en : - français - mathématiques	
		8 9 10 11 12 1 2 3 4 5 6	
	A venir : aides à l'évaluation en : - français - mathématiques	aides à l'évaluation en : - français - mathématiques	aides à l'évaluation en : - français - mathématiques
		8 9 10 11 12 1 2 3 4 5 6	8 9 10 11 12 1 2 3 4 5 6
VS		épreuves cantonales en : - français - mathématiques - allemand	épreuves cantonales en : - français - mathématiques - allemand
		8 9 10 11 12 1 2 3 4 5	8 9 10 11 12 1 2 3 4 5
			épreuve cantonale en : - L3
			8 9 10 11 12 1 2 3 4 5
VD		épreuves cantonales en : - français - mathématiques	épreuves d'établissement (certificat) en : - français - mathématiques - allemand (pour VSB et VSG) - anglais (pour VSB et VSG) - option (variable selon filières)
		8 9 10 11 12 1 2 3 4 5 6	8 9 10 11 12 1 2 3 4 5 6

Épreuve à visée sommative : prise en compte dans la promotion de l'élève
 Épreuve à visée diagnostique : résultats indicatifs
 Épreuve inscrite dans la loi scolaire du canton

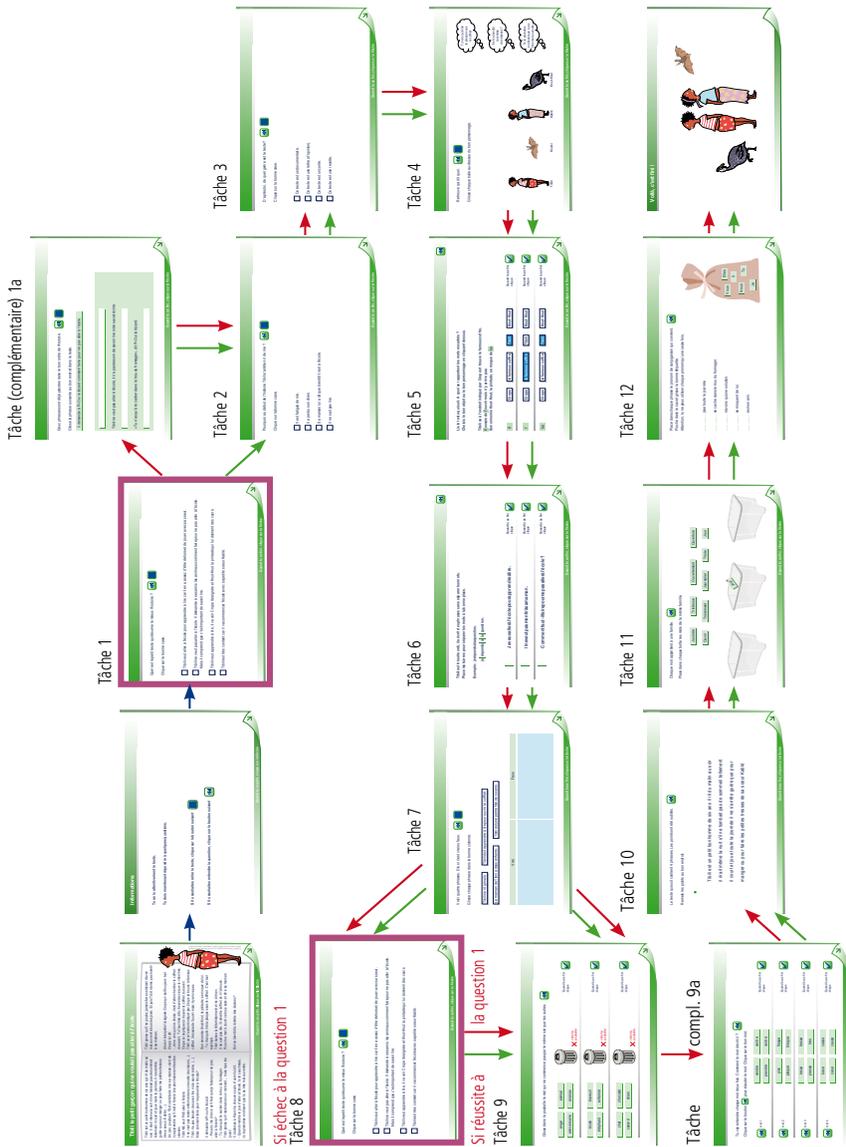
en jaune : mois de passation de l'épreuve
 8 9 10 11 12 | 1 2 3 4 5 6

8.3 Description de l'évaluation en deux phases en français

Une description plus complète présentant l'évaluation dans son ensemble et identifiant plus finement les attentes fondamentales concernées par les tâches est disponible sur le site de l'IRDPA sous www.irdp.ch/eprocom.

8.3.1 Description des tâches composant l'EA en français

Une description plus complète présentant l'évaluation dans son ensemble et identifiant plus finement les attentes fondamentales concernées par les tâches est disponible sur le site de l'IRDPA sous www.irdp.ch/eprocom.



8.3.2 Description des tâches composant l'EA en français

Une description plus complète présentant chaque tâche de l'EA est disponible sur le site de l'IRDP sous www.irdp.ch/eprocom.

IDENTIFICATION

Test de français

Je suis une fille: Je suis un garçon:

Ton prénom:

Classe de Madame Viridiana Merc

Date de naissance:

Quand tu es prêt, clique sur la flèche

APPROPRIATION DE L'OUTIL

Entraînement

Clique sur l'étoile:    

Donne l'ours à la petite fille:
(Clique sur l'ours et glisse-le au bon endroit)

- Clique sur la case A ✓
- Clique sur la case C ✓

A B C D 

Clique sur le bouton du son:  Règle le niveau sonore: 

Quand tu es prêt, clique sur la flèche

TEXTE DE RÉFÉRENCE

Tibili le petit garçon qui ne voulait pas aller à l'école

Tibili est un petit bonhomme de six ans qui rit du matin au soir. Il ritait même la nuit s'il ne tombait pas de sommeil tellement il court et joue toute la journée. Il ne s'arrête guère que pour manger ou pour faire les petites tresses de sa sœur, Kablé (...)

Un jour, pourtant, Tibili s'arrête de rire: sa maman vient de lui apprendre qu'il ira à l'école à la prochaine rentrée des classes.

Tibili NE veut PAS aller à l'école
Il ne veut pas rester enfermé dans une salle de classe (...)
Tibili n'a pas besoin de savoir lire ni de savoir écrire. (...)
Mais comment faire pour ne pas aller à l'école?

Il demande à Pi-ouï le lézard:
- Peux-tu me dire ce qu'il faut que je fasse pour ne pas aller à l'école?
- Tu n'as qu'à te cacher dans le trou du fromage.
Tibili pense qu'il restera bien un moment... mais tous les jours!

Il s'adresse à Koumi la cheuve-sours. Koumi lui dit:
- Quand viendra le jour d'aller à l'école, tu te coucheras, tu te plaindras en disant que tu es très mal au ventre.

Tibili pense qu'il ne pourra jamais faire semblant d'avoir mal au ventre tous les jours. Et puis Tibili n'aime pas mentir à sa maman.

Alors il interpelle l'araignée Crope qui sait toujours tout.
Crope lui dit:
- Je ne vois qu'une chose: c'est d'aller chercher le coffret du savoir. Tu fourreras et tu trouveras ce que tu cherches.
Crope lui explique où trouver le coffret du savoir.
Tibili va à l'endroit indiqué par Crope et trouve le fameux coffret. Il essaie de l'ouvrir mais n'y arrive pas.

Son amie Kruz-Kruz, la pintade, se moque de lui
- Tu n'as qu'à lire la plaque sous le coffret. C'est tout expliqué.
Tibili baisse la tête tristement et ne dit rien. Il ne sait pas lire. Il remet le coffret où il l'a trouvé.
Puis il se met à courir vers sa case et dit à sa maman:
Est-ce bientôt la rentrée des classes ?



Quand tu es prêt, clique sur la flèche

INDICATIONS COMPLÉMENTAIRES

Informations

Tu as lu attentivement le texte.

Tu dois maintenant répondre à quelques questions.

Durant le test, si tu souhaites relire le texte, clique sur le bouton suivant 

De même, tu auras la possibilité d'écouter la question en cliquant sur le bouton suivant 

Quand tu es prêt, clique sur la flèche

Tâche 1

Quel est le petit texte qui résume le mieux l'histoire?  

Clique sur la bonne case.

Tibili veut aller à l'école pour apprendre à lire car il en a assez d'être dehors et de jouer avec sa soeur.

Tibili ne veut pas aller à l'école. Il demande à ses amis les animaux comment faire pour ne pas aller à l'école. Mais il comprend que c'est important de savoir lire.

Tibili veut apprendre à lire, il va voir Crope l'araignée et Knut-Knut la pintade qui lui donnent des cours.

Tibili est très content car il va commencer l'école avec sa petite soeur Kablé.

Quand tu es prêt, clique sur la flèche 

Tâche complémentaire 1a

Deux phrases sont déjà placées dans le bon ordre de l'histoire.

Glisse la phrase suivante au bon endroit dans le texte.

Il demande à Pi-Oui le lézard comment faire pour ne pas aller à l'école.

Tibili ne veut pas aller à l'école; il n'a pas besoin de savoir lire ni de savoir écrire.

-Tu n'as qu'à te cacher dans le trou du fromager-, dit Pi-Oui le lézard.

Quand tu es fini, clique sur la flèche 

Tâche 2

Pourquoi au début de l'histoire Tibili s'arrête-t-il de rire ?  

Clique sur la bonne case.

Il est fatigué de rire.

Il a perdu son chien.

Sa maman lui a dit que bientôt il irait à l'école.

Il ne sait pas lire.

Quand tu es prêt, clique sur la flèche 

Tâche 3

D'après toi, de quel genre est le texte ?  

Clique sur la bonne case.

Ce texte est un documentaire.

Ce texte est une lettre (d'opinion).

Ce texte est un conte.

Ce texte est une recette.

Quand tu es prêt, clique sur la flèche 

Tâche 4

Retrouve qui dit quoi.

Glisse chaque bulle au-dessus du bon personnage.



Tibili Koumi Kablé Knut-Knut

Tu n'as qu'à lire la plaquette sous le coffre.

Est-ce bientôt la rentrée des classes ?

Tu te plaindras en disant que tu as très mal au ventre.

Quand tu as fini, clique sur la flèche

Tâche 5

Lis le texte qui suit. A quoi se rapportent les mots encadrés ?
Choisis le bon objet ou le bon personnage en cliquant dessus.

Tibili va à l'endroit indiqué par Crope et trouve le fameux coffre.
Il essaie de l'ouvrir mais n'y arrive pas.
Son ennemi Knut-Knut, la pintade, se moque de lui.

il	Crope	le fameux coffre	Tibili	Knut-Knut	Quand tu as fini clique	<input checked="" type="checkbox"/>
il	Crope	le fameux coffre	Tibili	Knut-Knut	Quand tu as fini clique	<input checked="" type="checkbox"/>
lui	Crope	le fameux coffre	Tibili	Knut-Knut	Quand tu as fini clique	<input checked="" type="checkbox"/>

Quand tu as fini, clique sur la flèche

Tâche 6

Tibili est très étourdi, il a écrit des phrases sans séparer les mots. 
Place les barres pour séparer les mots.

Exemple: Je réponds à la question.
Je | répons | à | la | question.

| Je | ve | ux | aller | à | l' | école | pour | app | ren | dre | à | lire | .
| Il | ne | ve | ut | pas | men | tir | à | sa | ma | man | .

Comment faut-il faire pour ne pas aller à l'école? Quand tu es fini clique sur 

Quand tu es prêt, clique sur la flèche 

Tâche 7

Voici quatre phrases. Dis si c'est vrai ou faux.  
Glisse chaque phrase dans la bonne colonne.

1)

2)

Vrai	Faux
<input type="text"/>	<input type="text"/>

3)

Vrai	Faux
<input type="text"/>	<input type="text"/>

Quand tu es prêt, clique sur la flèche 

Tâche 8

Quel est le petit texte qui résume le mieux l'histoire ?  

Clique sur la bonne case.

Tibili veut aller à l'école pour apprendre à lire car il en a assez d'être dehors et de jouer avec sa soeur.

Tibili ne veut pas aller à l'école. Il demande à ses amis les animaux comment faire pour ne pas aller à l'école. Mais il comprend que c'est important de savoir lire.

Tibili veut apprendre à lire, il va voir Crope l'araignée et Knut-Knut la pintade qui lui donnent des cours.

Tibili est très content car il va commencer l'école avec sa petite soeur Kablé.

Quand tu es prêt, clique sur la flèche 

Tâche 9

Glisse dans la poubelle le mot qui ne commence pas par le même son que les autres. 

ange animal 

anniversaire

école écureuil 

éléphant

chat chocolat 

 chien

Quand tu as fini clique sur 

Quand tu es prêt, clique sur la flèche 

Dans la capture d'écran ci-dessus, les mots qui ont été placés dans les poubelles sont :
1) ange; 2) enfermée; 3) canard.

Tâche complémentaire 9a

Tu vas entendre chaque mot deux fois. Comment le mot s'écrit-il? 

Clique sur le bouton  pour écouter le mot puis clique sur le mot que tu as entendu.

 mot 1

<input type="button" value="vendre"/>	<input type="button" value="vénir"/>
<input type="button" value="prendre"/>	<input type="button" value="vontre"/>

 mot 2

<input type="button" value="ploc"/>	<input type="button" value="flaque"/>
<input type="button" value="plaque"/>	<input type="button" value="blague"/>

 mot 3

<input type="button" value="treize"/>	<input type="button" value="tresse"/>
<input type="button" value="presse"/>	<input type="button" value="très"/>

 mot 4

<input type="button" value="base"/>	<input type="button" value="casse"/>
<input type="button" value="case"/>	<input type="button" value="cause"/>

Quand tu es prêt, clique sur la flèche 

Tâche 10

Le texte qui suit contient 4 phrases. Les points ont été oubliés. 

Remets les points au bon endroit.

•

Tibili est un petit bonhomme de six ans. Il rit du matin au soir,
Il rirait même la nuit s'il ne tombait pas de sommeil tellement
il court et joue toute la journée. Il ne s'arrête guère que pour
manger ou pour faire les petites tresses de sa soeur Kablé.

Tâche 11

Chaque mot appartient à une famille. 

Place dans chaque boîte les mots de la même famille.

Journées Tristesse Ouvertement Ouverture

Ouvrir Tristement Journalier Triste Jour



Quand tu as fini, clique sur la flèche 

Tâche 12

Place dans chaque phrase le pronom de conjugaison qui convient. 

Pioche dans le sac et glisse la bonne étiquette.

Attention, tu ne peux utiliser chaque pronom qu'une seule fois.

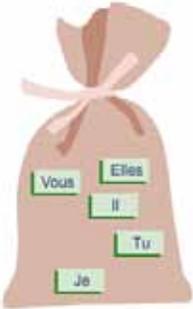
..... joue toute la journée.

..... se cache dans le trou du fromage.

Nous n'avons qu'une solution.

..... se moquent de lui.

..... es mon ami.



Quand tu es prêt, clique sur la flèche 

FIN



8.3.3 Description de l'épreuve papier (EP) en français

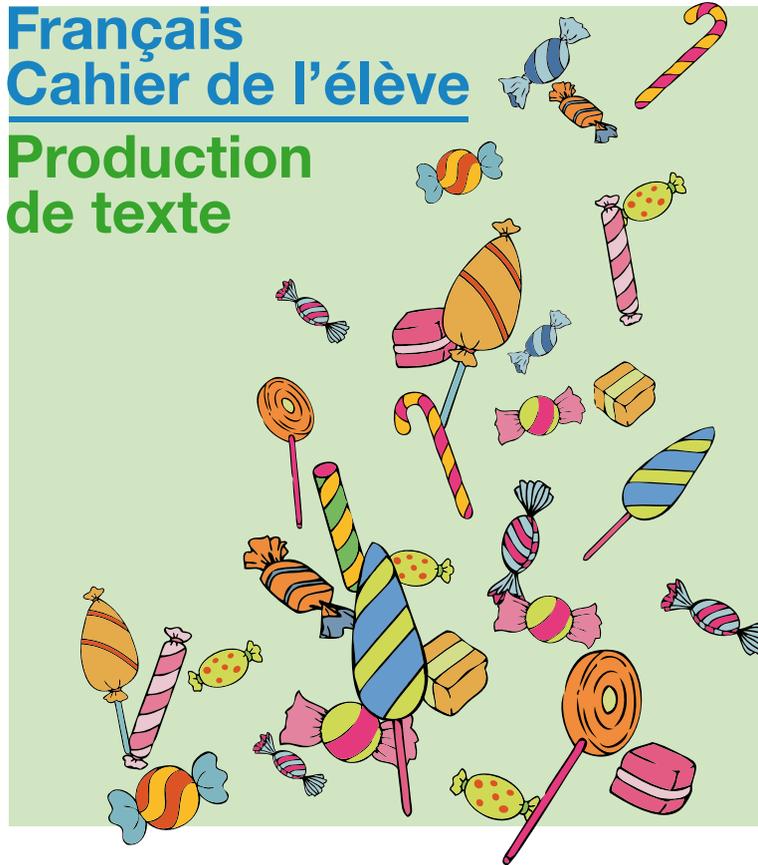
Une description plus complète présentant l'épreuve papier et identifiant plus finement les attentes fondamentales concernées par les tâches est disponible sur le site de l'IRDp sous www.irdp.ch/eprocom.

8.3.4 Description des tâches composant l'EP en français

Une description plus complète présentant chaque tâche de l'EP est disponible sur le site de l'IRDp sous www.irdp.ch/eocom.

Temps de production :

Temps de relecture :



Ta date de naissance : _____

Ton prénom : _____

Le nom de ton enseignant/e : _____

Langue(s) parlée(s) à la maison : _____

Fille: Garçon:

Compréhension

Sion, le 6 mai 2010

Chers camarades,

J'ai eu une grande discussion avec ma maman. Elle et moi, on n'a pas la même opinion en ce qui concerne les bonbons.

Elle ne veut pas que j'en mange du tout parce qu'elle trouve que c'est mauvais pour les enfants, alors que moi, je pense qu'on peut en manger.

Comme vous avez le même âge que moi, j'aimerais que vous me donniez votre avis.

Qu'est-ce que vous en pensez? Donnez-moi chacun votre opinion et deux arguments. Je suis impatient de recevoir vos lettres.

Je vous remercie par avance et je vous adresse, chers camarades, mes bonnes amitiés.

Paul

Questionnaire

Ta date de naissance: _____

Ton prénom: _____

Le nom de ton enseignant/e: _____

Lis la lettre de Paul, réponds aux questions et réalise les tâches ci-dessous.

1. Qui a écrit la lettre?
2. A qui cette lettre est-elle envoyée?
3. Dans la lettre, souligne en rouge l'opinion de la maman de Paul.
4. Dans la lettre, souligne en vert l'argument que donne la maman de Paul.
5. Cherche dans le texte ce qui introduit l'argument de la maman de Paul.
Mets une croix dans la case qui convient.
 - si;
 - parce que;
 - pour;
 - lorsque.
6. Dans la lettre, souligne en bleu l'opinion de Paul.
7. Pourquoi Paul a-t-il écrit cette lettre?
Mets une croix dans la case qui convient.
 - a) Paul a écrit cette lettre pour:
 - convaincre des camarades du même âge qu'il est bien pour les enfants de manger des bonbons;
 - donner l'occasion aux enfants qui l'ont reçue de la faire lire à leurs parents;
 - demander l'opinion d'autres enfants de son âge;
 - recevoir beaucoup de lettres en retour.
 - b) Souligne au crayon gris les mots qui t'ont permis de répondre.

Production

Sion, le 6 mai 2010

Chers camarades,

J'ai eu une grande discussion avec ma maman. Elle et moi, on n'a pas la même opinion en ce qui concerne les bonbons.

Elle ne veut pas que j'en mange du tout parce qu'elle trouve que c'est mauvais pour les enfants, alors que moi, je pense qu'on peut en manger.

Comme vous avez le même âge que moi, j'aimerais que vous me donniez votre avis.

Qu'est-ce que vous en pensez? Donnez-moi chacun votre opinion et deux arguments. Je suis impatient de recevoir vos lettres.

Je vous remercie par avance et je vous adresse, chers camarades, mes bonnes amitiés.

Paul

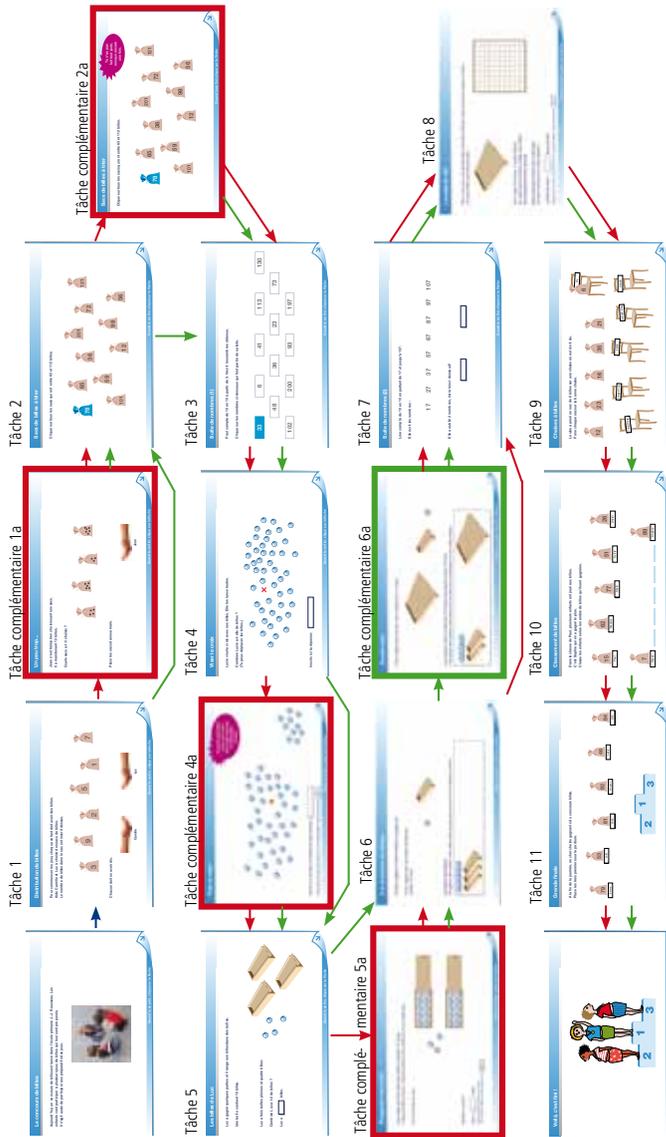
8.4 Description de l'évaluation en deux phases en mathématiques

*Une description plus complète présentant l'évaluation dans son ensemble et identifiant plus finement les attentes fondamentales concernées par les tâches est disponible sur le site de l'IRDp sous **www.irdp.ch/eprocom**.*

8.4.1 Description de l'épreuve adaptative (EA) en mathématiques

*Une description plus complète présentant l'épreuve adaptative et identifiant plus finement les attentes fondamentales concernées par les tâches est disponible sur le site de l'IRDp sous **www.irdp.ch/eprocom**.*

Chacune des tâches renvoie à une ou plusieurs attentes fondamentales. Le tableau des éléments de classification présente, pour chacune de ces attentes, le nombre de tâches les concernant.



8.4.2 Description des tâches composant l'EA en mathématiques

Une description plus complète présentant chaque tâche de l'EA est disponible sur le site de l'IRDp sous www.irdp.ch/eprocom.

IDENTIFICATION

Test de mathématiques

Je suis une fille: Je suis un garçon:

Ton prénom:

Classe de Madame Viridiana Marc

Date de naissance: Jour Mois

Quand tu es prêt, clique sur la flèche

APPROPRIATION DE L'OUTIL

Entraînement

Clique sur l'étoile:    

Donne l'ours à la petite fille:
(Clique sur l'ours et glisse-le au bon endroit)     

- Clique sur le 2 et le 4.
- Clique sur le 2. 

Écris le nombre 83 dans la case suivante:
(Clique dans la case avant d'écrire) 

Quand tu es prêt, clique sur la flèche

INTRODUCTION DU CONTEXTE : LE CONCOURS DE BILLES

Le concours de billes

Aujourd'hui, un concours de billes est lancé dans l'école primaire Rousseau. Les enfants vont participer à plusieurs jeux de billes qui leur sont proposés. Il s'agit aussi de participer aux préparatifs des jeux.



Quand tu es prêt, clique sur la flèche

Tâche 1

Distribution de billes

Pour commencer les jeux, chaque enfant doit avoir des billes.
Aide Camille et Luc à choisir des sacs de billes.
Le nombre de billes dans le sac est inscrit dessus.

Chacun doit en avoir dix.



Camille Luc

Quand tu es prêt, clique sur la flèche

Tâche complémentaire 1a

Un peu trop...

Jean s'est trompé en choisissant ses sacs.
Il a maintenant 13 billes.

Quels sacs a-t-il choisis?



Place les sacs dans sa main.



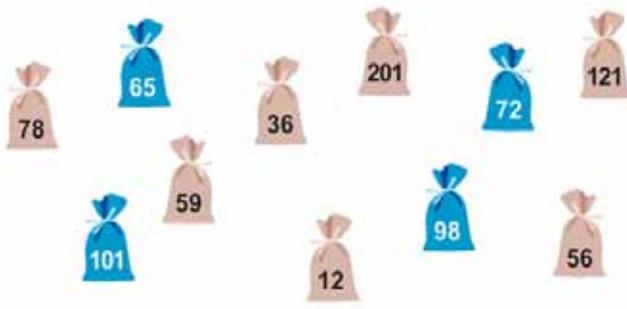
Jean

Quand tu es prêt, clique sur la flèche

Tâche 2

Sacs de billes à trier

Clique sur tous les sacs qui ont entre 63 et 112 billes.



Quand tu es prêt, clique sur la flèche

Tâche complémentaire 2a

Sacs de billes à trier

Clicke sur tous les sacs qui ont entre 63 et 112 billes.

Tu n'as pas fait tout juste, essaye encore une fois.

Quand tu as fini, clique sur la flèche

Tâche 3

Suite de nombres (1)

Paul compte de 10 en 10 à partir de 3. Il écrit les nombres obtenus.

Clicke sur les nombres ci-dessous qui font partie de sa liste.

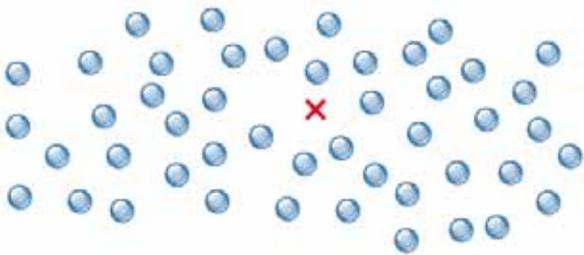
Quand tu as pris, clique sur la flèche

Tâche 4

Viser la croix

Lucie vise la croix avec ses billes. Elle les lance toutes

Combien Lucie a-t-elle lancé de billes ?
(Tu peux déplacer les billes.)



Inscris ici ta réponse:

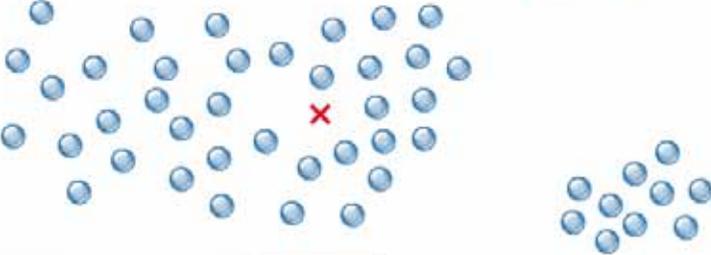
(Clique dans la case avant d'écrire. Si tu désires effacer, utilise cette touche )

Quand tu es prêt, clique sur la flèche 

Tâche complémentaire 4a

Viser la croix

Tu as fait une erreur.
Lucie veut t'aider:
elle a déjà regroupé
10 billes dans
un coin



Inscris ici le nombre de toutes les billes lancées:

(Clique dans la case avant d'écrire. Si tu désires effacer, utilise cette touche )

Quand tu es prêt, clique sur la flèche 

Tâche 5

Les billes de Luc

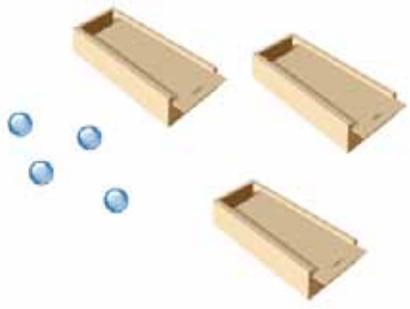
Luc a gagné quelques parties et il range ses billes dans des boîtes.

Une boîte contient 10 billes.

Luc a trois boîtes pleines et quatre billes:

Combien Luc a-t-il de billes?

Luc a billes.
(Clique dans la case avant d'écrire.
 Si tu désires effacer, utilise cette touche )



Quand tu es prêt, clique sur la flèche 

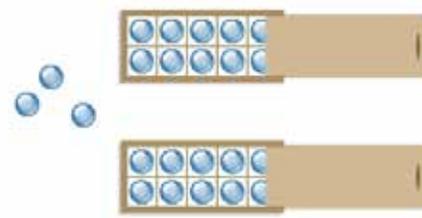
Tâche complémentaire 5a

Rangement des billes

Jean range ses billes dans des boîtes contenant 10 cases.

Il met une bille par case. Il réussit à remplir deux boîtes et il lui en reste trois.

Voici les billes de Jean:



Combien en a-t-il?

Jean a billes.
(Clique dans la case avant d'écrire.
 Si tu désires effacer, utilise cette touche )

Quand tu es prêt, clique sur la flèche 

Tâche 6

A la recherche des boîtes

Charles a gagné 37 billes et Marie en a gagné 42.

Ils les rangent dans des boîtes qui peuvent contenir 10 billes.



Les boîtes sont pleines.

Fais glisser des boîtes et des billes dans le cadre pour montrer les 37 billes de Charles :



Quand tu es prêt, clique sur la flèche

Tâche complémentaire 6a

Grande boîte

Elisa a gagné deux parties. Elle a désormais 127 billes.



Les grandes boîtes contiennent 100 billes et les petites boîtes 10 billes.

Fais glisser des boîtes et des billes dans le cadre pour représenter les 127 billes d'Elisa.



Quand tu es prêt, clique sur la flèche

Tâche 7

Suite de nombres (2)

Line compte de 10 en 10 en partant de 17 et jusqu'à 107.

Elle écrit les nombres :

17 27 37 57 67 87 97 107

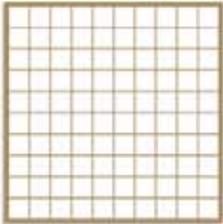
Elle a oublié 2 nombres, écris-les ci-dessous:
 (Clique dans la case avant d'écrire. Si tu désires effacer, utilise cette touche )

Quand tu es prêt, clique sur la flèche 

Tâche 8

La boîte de 100

Paul a reçu pour son anniversaire une boîte contenant cent cases pour ranger ses billes.
On peut mettre une bille par case.

Pour ranger ses billes, il remplit d'abord les cases de la première rangée, puis celles de la deuxième et ainsi de suite.

Après quelques parties, Paul a 68 billes.
Combien de lignes sont-elles pleines?

La boîte de Paul a lignes pleines.
 (Clique dans la case avant d'écrire. Si tu désires effacer, utilise cette touche )

Quand tu es prêt, clique sur la flèche 

Tâche 9

Chaises à billes

Louis a posé un sac de 6 billes sur une chaise où est écrit six.
Pose chaque sac sur la bonne chaise.

Quand tu es prêt, clique sur la flèche

Tâche 10

Classement de billes

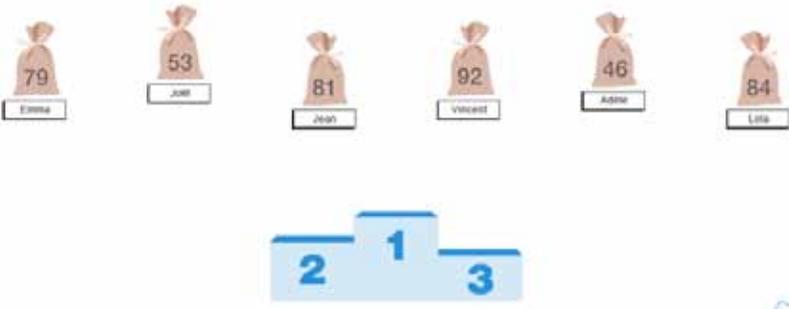
Dans la classe de Paul, plusieurs enfants ont joué aux billes.
C'est Sophie qui en a gagné le plus.
Classe les enfants selon le nombre de billes qu'ils ont gagnées.

Quand tu es prêt, clique sur la flèche

Tâche 11

Grande finale

A la fin de la journée, on cherche les gagnants du concours billes.
Place les trois premiers sur le podium.



Participant	Score
Emma	79
Joë	53
Jean	81
Vincent	92
Admè	46
Lita	84

Quand tu as fini, clique sur la flèche.

Voilà, c'est fini !



8.4.3 Description de l'épreuve papier (EP) en mathématiques

Une description plus complète présentant l'épreuve papier et identifiant plus finement les attentes fondamentales concernées par les tâches est disponible sur le site de l'IRDP sous **www.irdp.ch/eprocom**.

8.4.4 Description des tâches composant l'EP en mathématiques

Une description plus complète présentant chaque tâche est disponible sur le site de l'IRDP sous **www.irdp.ch/eprocom**.

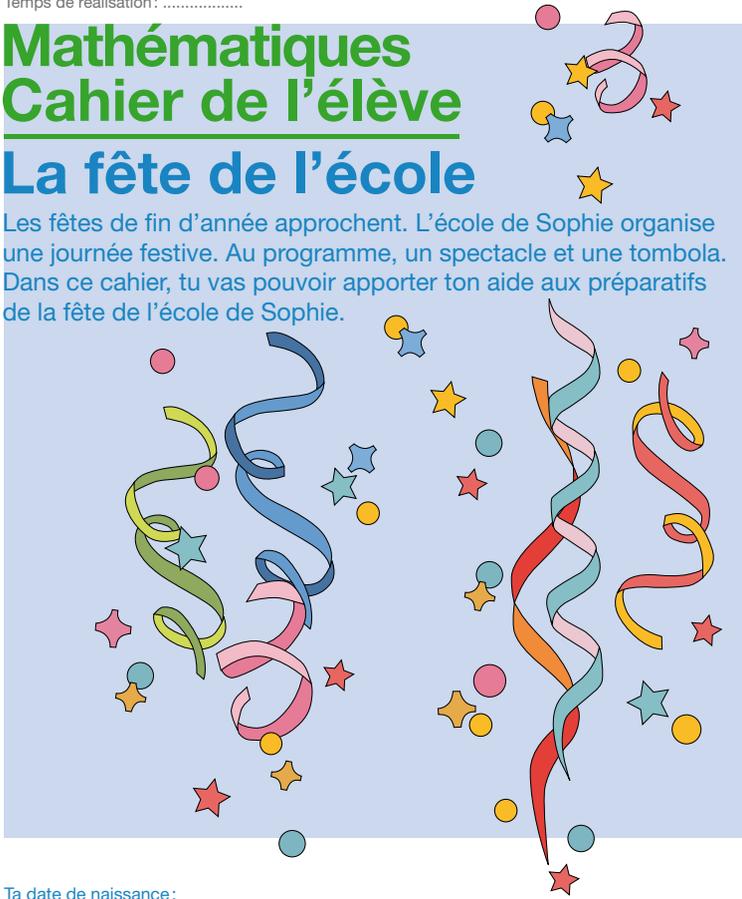
Temps de réalisation :

Mathématiques

Cahier de l'élève

La fête de l'école

Les fêtes de fin d'année approchent. L'école de Sophie organise une journée festive. Au programme, un spectacle et une tombola. Dans ce cahier, tu vas pouvoir apporter ton aide aux préparatifs de la fête de l'école de Sophie.



Ta date de naissance : _____

Ton prénom : _____

Le nom de ton enseignant/e : _____

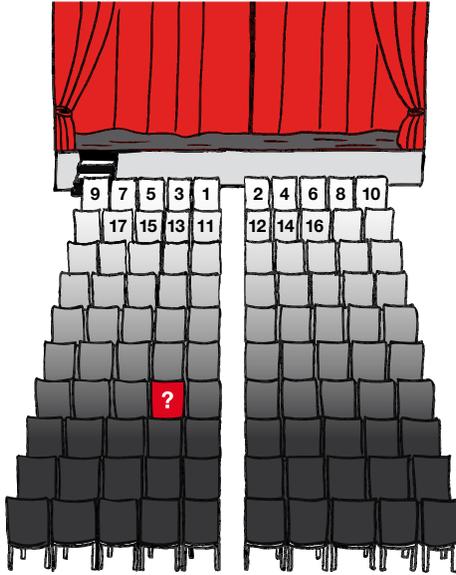
Langue(s) parlée(s) à la maison : _____

Fille : Garçon :

Le spectacle

**Pour le spectacle, on a commencé à numéroter les chaises.
Observe bien!**

Quel est le numéro de la chaise rouge?



Montre comment tu as fait:

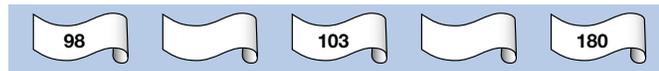
La tombola

Dans la tombola, on a mis les billets de 1 à 200.

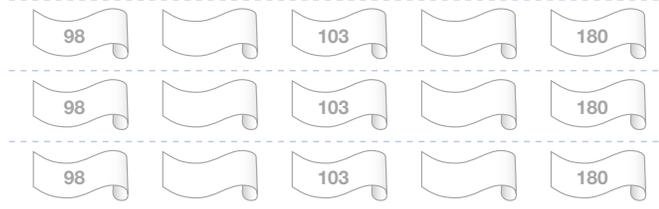
- a) Jean tire cinq billets. Il les classe du plus petit au plus grand.
Il a déjà posé les billets numérotés 98, 103 et 180.

Propose des nombres pour les billets blancs.

Ta réponse



Ici, tu peux faire des essais



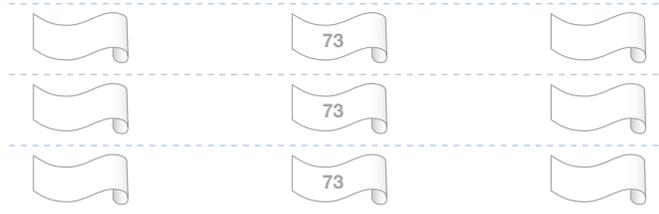
- b) Zoé tire trois billets. Elle voit que les chiffres 1, 3, 5, 7, 9 sont tous utilisés une seule fois. 7 et 3 sont déjà écrits sur un billet.

Propose des nombres pour les billets et classe-les du plus petit au plus grand.

Ta réponse



Ici, tu peux faire des essais



La tombola

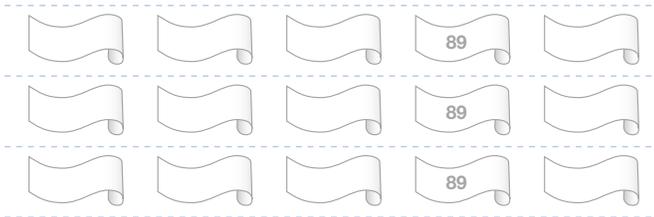
- c) Sophie tire cinq billets.
Elle voit que les chiffres 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 sont tous utilisés une seule fois.
8 et 9 sont déjà écrits.

Propose des nombres pour les billets et classe-les du plus petit au plus grand.

Ta réponse



Ici, tu peux faire des essais



Bibliographie

- Allal, L. (2002). Acquisition et évaluation des compétences en situation scolaire. In J. Dolz & E. Ollagnier (Eds), *L'énigme de la compétence en éducation* (pp. 75-94). Bruxelles: De Boeck Université.
- Allal, L. (2007). Régulation des apprentissages: orientation conceptuelle pour la recherche et la pratique en éducation. In L. Allal & L. Mottier Lopez (Eds), *Régulation des apprentissages en situation scolaire et en formation* (pp. 7-24). Bruxelles: De Boeck.
- Allal, L., Rouiller, Y., Saada-Robert, M. & Wegmuller, E. (1999). Gestion des connaissances orthographiques en situation de production textuelle. *Revue française de pédagogie*, 126, 53-69.
- Bain, D. (2000). De l'évaluation aux compétences: mise en perspective de pratiques émergentes. In J. Dolz & E. Ollagnier (Eds), *L'énigme de la compétence en éducation* (pp. 129-145). Bruxelles: De Boeck Université.
- Bain, D. (2010). Recourir à la généralisabilité pour évaluer des tests de maîtrise. *Mesure et évaluation en éducation*, 33(2), 35-63.
- Blais, J.-G. & Grondin, J. (2010). L'impact de la formulation des items dans les questionnaires d'enquête: une étude avec le modèle de Rasch pour les données polytomiques. *Mesure et évaluation en éducation*, 33(2), 95-126.
- Bloom, S. (Ed.). (1969). *Taxonomie des objectifs pédagogiques*. Montréal: Education nouvelle.
- Bronckart, J.-P. & Dolz, J. (1999). La notion de compétence : quelle pertinence pour l'étude de l'apprentissage des actions langagières?. In J. Dolz & E. Ollagnier (Eds), *L'énigme de la compétence en éducation* (pp. 27-44). Bruxelles: De Boeck Université.
- Cardinet, J. (1986). *Evaluation scolaire et mesure*. Bruxelles: De Boeck.
- Cardinet, J. & Laveault, D. (1996). *Dix années de travaux européens et nord-américains sur l'évaluation: quelles lignes de force?*. [S.l.] : [s.n.].
- Cèbe, S. & Goigoux, R. (2009). *Lector & lectrix: apprendre à comprendre les textes narratifs: CM1, CM2, 6e, Segpa*. Paris: Retz.
- Chartier, P., Vrignaud, P., Trosseille, B., Levasseur, J. & Tardieu, F. (2000). *Les banques de questions et les épreuves sur mesure: une nouvelle réponse aux besoins d'évaluation: 5e Biennale de l'éducation et de la formation*. Paris: INRP [Page Web]. Accès: <http://www.inrp.fr/biennale/5biennale/Contrib/394.htm> (Page consultée le 28 janvier 2013).
- Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP). (2010). *Plan d'études romand (PER)*. 1 coffret (5 fascicules). Neuchâtel: CIIP, Secrétariat général [Page Web]. Accès: <http://www.plandetudes.ch> (Page consultée le 28 janvier 2013).

- Crahay, M. (2006). Dangers, incertitudes et incomplétude de la logique de la compétence en éducation. *Revue française de pédagogie*, 154, 97-110.
- De Ketele, J.-M. (2006). Contrôles, examens et évaluation. In J. Beillerot & N. Mosconi (Dir.), *Traité des sciences et des pratiques de l'éducation* (pp.407-419). Paris: Dunod.
- Département de l'instruction publique, République et Canton de Genève (2000). *Les objectifs d'apprentissage de l'école primaire genevoise. Direction de l'enseignement primaire. Août 2000*. Genève: Département de l'instruction publique. (DIP, 2000).
- Dolz, J. & Schneuwly, B. (1996). Genres et progression en expression orale et écrite: éléments de réflexions à propos d'une expérience romande. *Enjeux*, 37/38, 49-75.
- Dubois, A. & Wirthner, M. (1990). *Saisir l'idée principale d'un texte: ambiguïté d'un objectif scolaire apparemment clair*. Neuchâtel: Institut de recherche et de documentation pédagogique (IRDP).
- Frey, K. (1986). European traditions of curriculum research. In U. Hameyer, K. Frey, H. Haft & F. Kuebart (Dir.), *Curriculum research in Europe* (pp. 11-16). Berwyn: Swets North America; Lisse: Swets & Zeitlinger.
- Gérard, F.-M. (2008). *Évaluer des compétences: guide pratique*. Bruxelles: De Boeck.
- Groupe de référence pour l'évaluation du travail des élèves (GRETEL). (2004). *Lignes directrices pour un système d'évaluation du travail des élèves dans la scolarité obligatoire*. Neuchâtel: Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin, Secrétariat général.
- Guignard, N. & Hirsig, F. (1989). DOMINONS la différenciation. *Math-Ecole*, 139, 13-32.
- Guignard, N. & Hirsig, F. (1991). La course autour du monde. *Math-Ecole*, 150, 11-21.
- Hadji, C. (1989). *L'évaluation, règles du jeu: des intentions aux outils*. Paris: ESF.
- Joannert, P. (2003). *Compétences et socioconstructivisme: un cadre théorique*. Bruxelles: De Boeck Université.
- Kempen, J.-L. van. (2008). *Pourquoi a-t-on développé les compétences à l'école?: pour mieux répondre aux exigences de l'économie ou pour favoriser une pédagogie plus efficace?*. Bruxelles: Union francophone des associations de parents de l'enseignement catholique (UFAPEC) [Page Web]. Accès: <http://www.ufapec.be/nos-analyses/pourquoi-a-t-on-developpe-les-competences-a-l-ecole/> (Page consultée le 28 janvier 2013).
- Lenoir, Y. (2011). Conséquences des conceptions curriculaires actuelles sur les modes évaluatifs. *Les dossiers des sciences de l'éducation*, 25, 13-40.
- Léonard, M. & Prigent, A. (2001). *Tibili le petit garçon qui ne voulait pas aller à l'école*. Paris: Magnard Jeunesse.
- Marc, V. & Wirthner, M. (2012). *Epreuves romandes communes: de l'analyse des épreuves cantonales à un modèle d'évaluation adapté au PER: rapport final du projet EpRoCom*. Neuchâtel: Institut de recherche et de documentation pédagogique (IRDP).

- Nidegger, Ch. (Ed.). (2011). *PISA 2009: compétences des jeunes romands: résultats de la quatrième enquête PISA auprès des élèves de 9^e année*. Neuchâtel: Institut de recherche et de documentation pédagogique (IRDP).
- Organisation de coopération et de développement économiques (OCDE). (2011). *Résultats du PISA 2009. Vol. 6: Élèves en ligne: technologies numériques et performance*. Paris: OCDE.
- Perrenoud, Ph. (1998). Construire des compétences: est-ce tourner le dos aux savoirs?. *Résonances*, 3, 3-7.
- Perrenoud, Ph. (2002). *Que faire de l'ambiguïté des programmes scolaires orientés vers les compétences?*. Genève: Université de Genève, Faculté de psychologie et des sciences de l'éducation.
- Pochon, L.-O. & Wirthner, M. (2007). Vers des épreuves de référence romandes: rapport final des rencontres romandes de 2006. Neuchâtel: Institut de recherche et de documentation pédagogique (IRDP).
- Rey, B., Carette, V., Defrance, A. & Kahn, S. (Réédition 2006). *Les compétences à l'école: apprentissage et évaluation*. Bruxelles: De Boeck.
- REY, B., Carette, V., Defrance, A. & Kahn, S. (2004). *Les compétences à l'école: apprentissage et évaluation*. Bruxelles: De Boeck.
- Rey, O. (2012). *Le défi de l'évaluation des compétences*. Lyon: ENS, IFé.
- Roegiers, X. (2004). *L'école et l'évaluation: des situations pour évaluer les compétences des élèves*. Bruxelles: De Boeck Université.
- Saporta, G. (1990). *Probabilités: analyse de données et statistiques*. Paris: Technip.
- Torres, J.-Ch. (2011). *L'évaluation dans les établissements scolaires: théories, objets et enjeux*. Paris: L'Harmattan
(chapitre 2: Les principes méthodologiques d'une évaluation des compétences).
- Vermot, B., Behrens, M. & Marc, V. (2011). Etude comparative de deux tests assistés par ordinateur en vue de l'élaboration d'épreuves communes en Suisse romande. In J.-G. Blais & J.-L. Gilles (Dir), *Evaluation des apprentissages et technologies de l'information et de la communication: le futur est à notre porte* (pp. 225-244). Québec: Presses de l'Université Laval.
- Watkins, C. (2010). Learning, performance and improvement. *Research matters*, 34, 16 p. [Page Web]. Accès: http://www.ioe.ac.uk/about/documents/Watkins_10_Lng_Perf_Imp_ev.pdf (Page consultée le 28 janvier 2013).

Glossaire

Compétence

Possibilité, pour un individu, de mobiliser un ensemble intégré de ressources en vue d'exercer efficacement une activité considérée généralement comme complexe. (PER, 2010).

Une compétence est la possibilité, pour un élève, de mobiliser un ensemble de savoirs, savoir-faire et de savoir-être pour résoudre des situations. (Roegiers, 2000).

Une authentique compétence est la capacité à répondre à des situations complexes et inédites par une combinaison nouvelle de procédures connues; et non pas seulement à répondre par une procédure stéréotypée à un signal préétabli. (Rey, 2006).

Évaluation

... à visée sommative

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.

... à visée certificative

évaluation sommative qui conduit à une attestation officielle de compétences globales au terme d'une période de formation et peut contribuer à la promotion de l'élève.

... de sélection ou d'orientation

évaluation certificative permettant de pronostiquer, chez les élèves, des niveaux de compétences, certains dépassant les exigences définies pour la période de formation qu'ils achèvent; le but est de répartir les élèves dans les niveaux ou les filières subséquents.

... à visée diagnostique

elle 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.

... bilan

évaluation scientifique qui doit permettre de recueillir, à des moments-clés de la scolarité, des données représentatives de la maîtrise de tous les objectifs du plan d'études et de fournir des éléments permettant de les analyser et de les comprendre de manière à pouvoir permettre le pilotage du système éducatif par les responsables scolaires. Par conséquent, l'évaluation bilan ne peut fournir d'information sur les élèves pris individuellement.

Item

Élément d'une évaluation permettant d'extraire une information, souvent considéré comme unité dans l'attribution de points. Le terme est fréquemment associé (voire confondu) à une question lorsque celle-ci est courte et fermée ou à choix multiple.

Lorsqu'une question est ouverte, nous prenons le parti de considérer qu'elle contiendra généralement plusieurs items.

Procédure

Nous réservons le terme procédure à la démarche de résolution d'un problème ou à la réalisation d'une tâche complexe par l'élève. Cela nous permet d'une part, de repérer des procédures connues (ou censées l'être) et des procédures inédites et, d'autre part, de caractériser les réponses des élèves en termes de graduation des procédures engagées, expertes, semi-expertes ou exploratoires.

Ressources

Nous retenons ici la définition que propose Roegiers (2004): « Les **ressources** désignent l'ensemble des savoirs, savoir-faire, savoir-être, savoirs d'expérience, ... que l'apprenant mobilise pour résoudre une situation »

Savoirs

Le terme **savoir** comprend les contenus et les connaissances; l'identification des savoirs en jeu dans une évaluation scolaire se réfère essentiellement au programme d'études, le PER en l'occurrence. Les savoirs font partie des ressources à mobiliser en situation.

Savoir-faire

Le terme **savoir-faire** comprend les activités qui touchent à un savoir; il s'agit donc autant d'application de règles, de procédures automatisées, de gestes précis et entraînés, ... Les savoir-faire font partie des ressources à mobiliser en situation.

Situation cible

Il s'agit d'une situation pour laquelle l'élève ne possède pas de procédures automatisées lui permettant sa résolution. Une **situation inédite et complexe** implique que l'élève choisisse et combine les ressources dont il dispose.

Tâche complexe

Dans le cadre de notre projet, une **tâche complexe** s'inscrit dans une situation d'évaluation, donc dans une épreuve; elle présente une situation complexe, si possible inédite, choisie en lien avec une discipline et des éléments du référentiel. Il est possible d'étendre le terme à des champs disciplinaires, c'est-à-dire faisant appel à plusieurs disciplines (p. ex. sur des questions liées à l'éducation en vue du développement durable).

Tâche spécifique

Dans le cadre de notre projet, une **tâche spécifique** s'inscrit dans une situation d'évaluation, donc dans une épreuve et cible un nombre déterminé d'éléments du référentiel; il s'agit souvent de questions simples et fermées pour lesquelles on attend que l'élève exprime ses connaissances sans avoir à les combiner.

Tests de référence

On désigne sous **tests de référence** les évaluations système menées au niveau suisse sur la base des « compétences fondamentales » (standards nationaux de formation).



Dans l'article 15 de la Convention scolaire romande (CSR), la CIIP (Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin) exprime sa volonté de développer des épreuves romandes communes. Le contenu de cet article soulève toutefois la question des finalités de telles épreuves. Le présent rapport, sans y apporter directement une réponse, met en avant les différentes réflexions menées parallèlement au projet EpRoCom (Épreuves romandes communes) et dont le but premier est de caractériser une évaluation permettant la mesure des objectifs du PER (Plan d'études romand).

Nous présentons tout d'abord les situations cantonales en matière d'évaluation ainsi que le cadre du mandat. Nous nous penchons ensuite sur des éléments théoriques permettant d'étayer les choix effectués dans nos développements ainsi que sur un modèle évaluatif susceptible de répondre aux exigences d'une évaluation adaptée au PER.

Enfin, nous décrivons une épreuve-prototype diagnostique (en mathématiques et en français) permettant une première mise à l'essai du modèle évaluatif proposé. Ceci nous amène à envisager des éléments de classification des items et des tâches complexes, notamment dans la perspective de développer une banque d'items, ainsi que le traitement possible des données et les améliorations souhaitées.