

CEDRE

Cycle des Évaluations Disciplinaires Réalisées sur Échantillons

Rapport technique

Anglais 2016

École

Auteurs :
Sylvie BEUZON
Marion LE CAM
Corinne MARCHOIS
Louis-Marie NINNIN
Thierry ROCHER
Ronan VOURC'H

Bureau de l'évaluation des élèves
DEPP - Direction de l'évaluation, de la prospective et de la performance
Ministère de l'éducation nationale

Mars 2018

Table des matières

Introduction	3
1 Cadre d'évaluation	4
1.1 Objectifs	4
1.2 Connaissances et compétences visées	5
1.3 Construction du test	9
1.4 Passation des évaluations	11
2 Sondage	12
2.1 Méthodes	12
2.2 Echantillonnage	16
2.3 Etat des lieux de la non-réponse	18
2.4 Redressement	20
2.5 Précision	22
3 Analyse des items	25
3.1 Méthodologie	25
3.2 Codage des réponses aux items	28
3.3 Résultats	32
4 Modélisation	33
4.1 Méthodologie	33
4.2 Résultats	39
4.3 Calcul des scores	44
5 Construction de l'échelle	45
5.1 Méthode	45
5.2 Caractérisation des groupes de niveaux	46
5.3 Exemples d'items	49
6 Variables contextuelles et non cognitives	62
6.1 Variables sociodémographiques et indice de position sociale	62
6.2 Élaboration des questionnaires de contexte	65
6.3 Motivation des élèves face à la situation d'évaluation	66
7 Annexe	68
Références	71

Introduction

La DEPP met en place des dispositifs d'évaluation des acquis des élèves reposant sur des épreuves standardisées. Elle est également maître d'œuvre pour la France des évaluations internationales telles que PIRLS ou PISA. Ces programmes d'évaluations sont des outils d'observation des acquis des élèves pour le pilotage d'ensemble du système éducatif (Trosseille & Rocher, 2015). Les évaluations du CEDRE (Cycle d'Évaluations Disciplinaires Réalisées sur Échantillons) révèlent ainsi, en référence aux programmes scolaires, les objectifs atteints et ceux qui ne le sont pas. Ces évaluations doivent permettre d'agir au niveau national sur les programmes des disciplines, sur l'organisation des apprentissages, sur les contextes de l'enseignement, sur des populations caractérisées.

Leur méthodologie de construction s'appuie sur les méthodes de la mesure en éducation et sur des modélisations psychométriques. Ces évaluations concernent de larges échantillons représentatifs d'établissements, de classes et d'élèves. Elles permettent d'établir des comparaisons temporelles afin de suivre l'évolution des performances du système éducatif.

Ce rapport présente l'ensemble des méthodes qui sont employées pour réaliser les évaluations du cycle CEDRE, en balayant des aspects aussi divers que la construction des épreuves, la sélection des échantillons ou bien la modélisation des résultats. L'objectif est de rendre accessible les fondements méthodologiques de ces évaluations, dans un souci de transparence. La publication de ce rapport fait d'ailleurs partie des engagements pris par la DEPP dans le cadre du processus de certification des évaluations du cycle CEDRE.

1 Cadre d'évaluation

1.1 Objectifs

Le cycle des évaluations disciplinaires réalisées sur échantillons (CEDRE) établit des bilans nationaux des acquis des élèves en fin d'école et en fin de collège. Il couvre les compétences des élèves dans la plupart des domaines disciplinaires en référence aux programmes scolaires. La présentation des résultats permet de situer les performances des élèves sur des échelles de niveau allant de la maîtrise pratiquement complète de ces compétences à une maîtrise bien moins assurée, voire très faible, de celles-ci. Renouvelées tous les six ans (tous les cinq ans à partir de 2012), ces évaluations permettent de répondre à la question de l'évolution du niveau des élèves au fil du temps.

Ces évaluations n'ont pas valeur de délivrance de diplômes, ni d'examen de passage ou d'attestation de niveau ; elles donnent une photographie instantanée de ce que savent et savent faire les élèves à la fin d'un cursus scolaire. En ce sens, il s'agit bien d'un bilan. Destinées à être renouvelées périodiquement, ces évaluations-bilans permettent également de disposer d'un suivi de l'évolution des acquis des élèves dans le temps. Pour cette raison, les épreuves ne peuvent pas être totalement rendues publiques car, devant être en grande partie reprises lors des prochains cycles d'évaluation, elles ne doivent pas servir d'exercices dans les classes.

Ces évaluations apportent un éclairage qui intéresse tous les niveaux du système éducatif, des décideurs aux enseignants sur le terrain, en passant par les formateurs d'enseignants : elles informent sur les compétences et les connaissances des élèves à la fin d'un cursus ; elles éclairent sur l'attitude et la représentation des élèves à l'égard de la discipline ; elles interrogent les pratiques d'enseignement au regard des programmes ; elles contribuent à enrichir la réflexion générale sur l'efficacité et la performance de notre système éducatif.

Ces évaluations étant passées auprès d'échantillons statistiquement représentatifs de la population scolaire de France métropolitaine, aucun résultat par élève, établissement ni même par département ou académie ne peut être calculé.

CEDRE a débuté en 2003 avec l'évaluation des compétences générales. Afin d'assurer une comparabilité dans le temps, l'évaluation est reprise pour chaque discipline selon un cycle de six ans jusqu'en 2012, et de cinq ans depuis 2012 (tableau 1).

Tableau 1 – Cycle des évaluations disciplinaires réalisées sur échantillons (CEDRE) depuis 2003

Discipline évaluée	Début du cycle	Reprises	
Maîtrise de la langue et compétences générales	2003	2009	2015
Langues étrangères	2004	2010	2016
Attitude à l'égard de la vie en société	2005	–	–
Histoire, géographie et éducation civique	2006	2012	2017
Sciences	2007	2013	2018
Mathématiques	2008	2014	2019

1.2 Connaissances et compétences visées

1.2.1 Constitution des épreuves

En anglais, l'évaluation a été proposée dans quatre activités langagières : la compréhension de l'oral, la compréhension de l'écrit, l'expression écrite et, pour la première fois en 2016, l'expression orale en continu. Cette évaluation a été élaborée à partir des objectifs fixés par les programmes officiels entrés en vigueur à la rentrée 2008 (BO Hors-série n°3 du 19 juin 2008), programmes adossés au Cadre européen commun de référence pour les langues (CECRL) (voir encadré ci-dessous).

En fin de CM2, les élèves doivent avoir acquis les compétences nécessaires à la communication élémentaire définie par le niveau A1 du Cadre européen commun de référence pour les langues qui constitue par ailleurs la référence fondamentale pour l'enseignement, les apprentissages et l'évaluation des acquis en langues vivantes. [...] les activités orales de compréhension et d'expression sont une priorité. Le vocabulaire s'enrichit et les composantes sonores de la langue restent une préoccupation constante : accentuation, mélodies, rythmes propres à la langue apprise. En grammaire, l'objectif visé est celui de l'utilisation de formes élémentaires : phrase simple et conjonctions de coordination. L'orthographe des mots utilisés est apprise. Les connaissances sur les modes de vie du pays viennent favoriser la compréhension d'autres façons d'être et d'agir.

A la fin du CM2 les élèves doivent être capables de :

- communiquer : se présenter, répondre à des questions et en poser ;

- **comprendre des consignes, des mots familiers et des expressions très courantes ;**
- **connaître quelques éléments culturels d'un autre pays.**

Conformément aux préconisations des programmes, les situations d'évaluation relèvent pour la plupart d'entre elles du niveau A1. Néanmoins, afin d'apprécier au mieux les différentes performances des élèves, des items de niveau A1+ voire A2 ont également été proposés.

1.2.2 Objectifs d'évaluation et supports

Dans chacune des activités langagières retenues, on a évalué les connaissances et compétences des élèves. Pour ce faire, on a retenu plusieurs grands domaines de savoirs et de savoir-faire, gradués en fonction de la complexité croissante des opérations mentales nécessaires pour les mettre en œuvre.

En **compréhension de l'oral**, on vérifie que les élèves sont capables, dans un message sonore, de percevoir et reconnaître des phonèmes spécifiques dans un mot et de repérer une syllabe accentuée. On a vérifié qu'ils connaissent des termes lexicaux de base, appartenant aux programmes en vigueur ainsi que des blocs lexicalisés. Enfin, on a évalué leur capacité à dégager les principales informations d'un document sonore : comprendre une consigne simple, un message oral bref, une brève description de quelqu'un, d'un lieu, d'un objet ; une histoire brève et un court dialogue. Pour la première fois, des supports vidéo ont également été proposés.

En **compréhension de l'écrit**, on mesure les aptitudes des élèves à reconnaître dans un support des expressions mémorisées, des consignes de classe et un lexique de la vie quotidienne ; à se faire une idée du contenu d'un texte accompagné ou non d'une aide visuelle et à identifier un type de document. On a évalué leur capacité à dégager les principales informations dans un document écrit : texte informatif, lettre, carte postale, courriel, SMS, questionnaire, texte narratif et texte descriptif (brève description d'un lieu, d'une personne, d'un objet). Enfin, a été mesurée leur capacité à synthétiser l'information, comme par exemple choisir le résumé correspondant à un texte.

En **expression écrite**, on vérifie que les élèves sont capables de copier des mots isolés et des textes courts ; de remettre les mots d'une phrase dans l'ordre ; d'écrire un message électronique simple, ou une courte carte postale, en référence à des modèles ; de produire de manière autonome quelques phrases et de répondre à un questionnaire.

Enfin, en **expression orale en continu**, on évalue les aptitudes des élèves à se présenter et parler de leur environnement immédiat à partir de supports

Tableau 2 – Définition des compétences en compréhension de l’oral (évaluation 2016)

Compétences	Objectifs d'évaluation
Reconnaître / Percevoir	1. Identifier 2. Repérer un phonème spécifique dans un mot, repérer la syllabe accentuée
Connaître et reconnaître	3. Connaître les termes lexicaux de base. 4. Connaître les blocs lexicalisés
Dégager les principales informations d'un document sonore	5. Comprendre une consigne simple 6. Comprendre globalement un message oral bref 7a Comprendre quelqu'un qui se présente 7b Comprendre une brève description de quelqu'un, d'un lieu, d'un objet avec localisation 8. Comprendre une histoire brève ; être capable de recueillir des informations dans un texte de 20 mots environ 9. Comprendre un court dialogue

iconographiques variés et déclencheurs de parole. Pour des raisons de faisabilité, seul un sous-échantillon d'élèves (2 ou 3 élèves par classe) a été évalué dans cette activité langagière. L'élaboration des grilles de compétences (tableaux 2, 3, 4 et 5) en vue de la construction des items s'est fondée sur les documents de référence : les programmes officiels (BO Hors-série n°3 du 19 juin 2008).

Tableau 3 – Définition des compétences en compréhension de l'écrit (évaluation 2016)

Compétences	Objectifs d'évaluation
Connaître et reconnaître	1. le lexique de la vie quotidienne. 2a les expressions figées mémorisées 2b les courtes phrases 2c les questions et les réponses associées 3. les consignes de classe
Comprendre des textes courts et simples	
Se faire une idée du contenu d'un texte	4. Se faire une idée du contenu d'un texte accompagné d'une aide visuelle 5. Se faire une idée du contenu d'un texte sans aide visuelle 6. Identifier le type de document
Dégager les principales informations	7. d'un texte informatif 8. d'une lettre, d'une carte postale, d'un courriel, d'un SMS, d'un "chat" 9. d'un questionnaire 10. d'un texte narratif 11. d'un texte descriptif (brève description d'un lieu, d'une personne, d'un objet)
Synthétiser l'information	12. Choisir le résumé correspondant au texte

Tableau 4 – Définition des compétences en expression écrite (évaluation 2016)

Compétences	Objectifs d'évaluation
Etre capable de	1a. Copier des mots isolés et des textes courts 1b. Remettre les mots d'une phrase dans l'ordre 2. Ecrire un message électronique simple, ou une courte carte postale, en référence à des modèles 3a. Produire de manière autonome quelques phrases 3b. Répondre à un questionnaire

Tableau 5 – Définition des compétences en expression orale en continu (évaluation 2016)

Objectifs d'évaluation	
1. Se présenter	2. Parler de son environnement immédiat

1.3 Construction du test

Le bureau de l'évaluation des élèves de la DEPP élabore des évaluations par disciplines et niveaux scolaires. La préparation des unités et de leurs constituants fait intervenir des concepteurs, généralement des enseignants. La coordination est assurée par un chef de projet, membre de l'équipe du bureau de l'évaluation des élèves. Une application dédiée leur permet de créer, modifier ou éditer leur unité ; en outre cette application permet au chargé d'étude de gérer l'ensemble de l'évaluation (cf. plus loin l'encadré « GEODE »).

1.3.1 Elaboration des items

Les items sont le fruit d'un travail collectif des concepteurs, encadré par le chef de projet, l'inspection et l'inspection générale. Un item proposé par un concepteur, pédagogue de terrain ayant une bonne connaissance des pratiques de classe, fait l'objet d'une discussion contradictoire jusqu'à aboutir à un consensus. L'item est alors soumis à un « cobayage », c'est-à-dire une passation auprès d'une ou plusieurs classes pour estimer sa difficulté et recueillir les réactions des élèves.

Un équilibre de proportion entre les items considérés comme étant de difficulté « facile », « moyenne » ou « difficile » est recherché. Pour les quatre domaines évalués, la compréhension de l'écrit, la compréhension de l'oral, l'étude de la langue et la production d'écrit, certains items sont identiques à ceux proposés en 2010 et 2004 afin d'assurer une comparabilité de qualité.

Deux types de formats de questions sont utilisés : les questions fermées (QCM, QCM-images, série, série-images) et les questions ouvertes appelant une réponse écrite (réponse courte - un chiffre, un nombre - ou réponse longue - production en autonomie de l'élève). Un entraînement est prévu au début de chaque cahier afin de familiariser les élèves avec le type de question rencontré.

Les réponses des formats QCM, QCM-images sont saisies de manière automatisée à la fin de la passation. Les réponses des formats série et série-images sont saisies de manière automatisée et donnent lieu à un regroupement ultérieur de leurs propositions. Dans le cas de ces séries, des seuils statistiques ont été établis pour valider les réponses des élèves. Les réponses des formats " réponse libre de l'élève " sont corrigées par des experts via une interface Internet. Ce dispositif de correction à distance s'appuie sur le logiciel AGATE (cf. partie « Analyse des items »).

GEODE (Gestion électronique d'outils et documents d'évaluation) : un outil de création et de stockage des évaluations**Objectifs**

Le bureau de l'évaluation des élèves coordonne chaque année plusieurs évaluations afin d'apprécier le niveau de connaissances et de compétences des élèves en référence aux programmes officiels. Ces évaluations utilisent des livrets d'évaluation sur format papier et/ou électroniques.

L'application GEODE (gestion électronique d'outils et documents d'évaluation) est une application de création et de gestion dématérialisées des évaluations. Développée en 2009, elle a pour objectif de soutenir de bout en bout le processus de création des exercices et de constitution des cahiers et supports électroniques, allant jusqu'au bon à imprimer pour les évaluations papiers ou la génération d'une maquette de site web pour l'évaluation électronique.

L'application permet la conservation, l'indexation et la recherche des documents ou fichiers joints. Une partie des données textuelles, images, sons ou vidéos y est donc stockée que ce soit pour les évaluations papiers (cahier d'évaluations) ou les évaluations électroniques (outil de maquettage).

Principes fonctionnels

GEODE permet ainsi l'harmonisation des pratiques et formats de documents. La dématérialisation des documents rend indépendant l'éditeur (OpenOffice, Word,...) tout en permettant des variantes selon les disciplines. L'application dispose d'une GED (gestion électronique de documents) intégrée capable de gérer du texte, des images, du son et de la vidéo sous forme d'objets. Les cahiers sont générés au format Open Office principalement pour le format « papier », l'utilisation de la même technologie permet de générer du HTML pour la partie évaluation électronique (outil de maquettage).

1.3.2 Constitution des cahiers

Dans le cadre d'une évaluation sur support papier, le test se compose d'un ensemble de cahiers, constitués de blocs, qui sont eux-mêmes composés d'unités (ensemble d'items). Pour l'évaluation CEDRE anglais école 2016, 13 blocs de compréhension de l'écrit et d'expression écrite sont constitués et répartis au sein de 13 cahiers. Pour la compréhension de l'oral, 6 blocs sont constitués et répartis dans 6 autres cahiers. L'évaluation CEDRE anglais école 2016 est

constituée de 59 items de 2010 (représentant 30 % du nombre total d'items de l'évaluation) repris à l'identique afin d'assurer la comparabilité dans le temps et de 197 nouveaux items qui ont fait l'objet d'une expérimentation en 2015.

Afin de pouvoir évaluer un nombre important d'items sans allonger le temps de passation pour l'élève, CEDRE utilise la méthodologie des cahiers tournants. Les items sont ainsi répartis dans des blocs d'une durée de 12 minutes et les blocs sont ensuite distribués dans les cahiers tout en respectant certaines contraintes : chaque bloc doit se retrouver un même nombre de fois au total et chaque association de blocs doit figurer au moins une fois dans un cahier. Ce dispositif, couramment utilisé dans les évaluations bilans, notamment les évaluations internationales, permet d'estimer la probabilité de réussite de chaque élève à chaque item sans que chaque élève ait à passer l'ensemble de ceux-ci.

Au final, pour l'évaluation CEDRE anglais école 2016, chaque cahier comprend deux séquences de 45 minutes au cours desquelles l'élève est tout d'abord évalué en compréhension de l'oral (15 minutes), puis en compréhension de l'écrit et expression écrite (30 minutes). Les séances se terminent par un questionnaire de contexte, d'une durée de 15 minutes environ (une première partie en fin de séance 1 et une deuxième partie en fin de séance 2), identique dans tous les cahiers, dans lequel l'élève doit répondre à des questions concernant notamment l'environnement familial dans lequel il évolue, ses projets scolaires et professionnels, sa perception de la matière et de son environnement scolaire.

1.4 Passation des évaluations

La passation de l'évaluation finale a eu lieu en mai 2016. Comme en 2010, cette évaluation a été précédée d'une expérimentation l'année N- 1 de façon à tester un grand nombre d'items auprès d'un échantillon réduit d'établissements. Dans chaque établissement, une personne a été désignée comme étant l'administrateur du test, son rôle étant de veiller au strict respect de la procédure à suivre pour que l'évaluation soit passée dans les meilleures conditions, quel que soit l'établissement ; la collecte de l'information s'est faite par questionnaires "papier-crayon".

L'anonymat des élèves et des personnels est respecté, chaque cahier étant repéré par un numéro. Une fois l'évaluation terminée, les cahiers et questionnaires sont renvoyés dans des conditionnements prévus à cet effet, préaffranchis et pré-étiquetés. Aucun travail de correction n'a été demandé aux établissements.

2 Sondage

2.1 Méthodes

2.1.1 Sondage par grappes stratifié

Dans le premier degré, nous ne disposons pas des informations auxiliaires présentes dans les bases de sondage de la DEPP, telle que la PCS des parents par exemple. Il n'est donc pas possible de réaliser un tirage équilibré comme c'est le cas pour les évaluations CEDRE en 3e.

Le tirage consiste donc simplement en un sondage par grappes stratifié. La stratification porte généralement sur la zone de scolarisation et tous les élèves de CM2 des écoles sélectionnés participent. Le choix de sondages par grappe est motivé par la facilité de gestion. En effet, le fait de sélectionner tous les élèves d'une école permet d'éviter de mettre en place des procédures de tirage au sort d'élèves une fois les écoles tirées.

Par ailleurs, au moment du tirage de l'échantillon, les écoles ayant déjà été sélectionnée pour une autre évaluation la même année sont exclues de la base de sondage. Les probabilités d'inclusion sont donc recalculées pour tenir compte de ces exclusions tout en gardant une représentativité nationale (cf. encadré « tirage après élimination de la base des échantillons précédemment tirés »).

2.1.2 Redressement de la non réponse : calage sur marges

Comme toute enquête réalisée par sondage, les évaluations des élèves sont exposées à la non-réponse. Bien que les taux de retour soient élevés, il est nécessaire de tenir compte de la non-réponse dans les estimations car celle-ci n'est pas purement aléatoire (par exemple, la non-réponse est plus élevée chez les élèves en retard). Afin de la prendre en compte, un calage sur marges est effectué à l'aide de la macro CALMAR, également disponible sur le site Internet de l'INSEE. La méthode de calage sur marges consiste à modifier les poids de sondage d_i des répondants de manière à ce que l'échantillon ainsi repondéré soit représentatif de certaines variables auxiliaires dont on connaît les totaux sur la population (Sautory, 1993). C'est une méthode qui permet de corriger la non-réponse mais également d'améliorer la précision des estimateurs. En outre, elle a pour avantage de rendre cohérents les résultats observés sur l'échantillon pour ce qui concerne des informations connues sur l'ensemble de la population.

Les nouveaux poids w_i , calculés sur l'échantillon des répondants S' , vérifient l'équation suivante pour les K variables auxiliaires sur lesquelles porte le calage :

$$\forall k = 1 \dots K, \sum_{i \in S'} w_i X_i^k = \sum_{i \in U} X_i^k \quad (1)$$

Ils sont obtenus par minimisation de l'expression $\sum_{i \in S'} d_i G(\frac{w_i}{d_i})$ où G désigne une fonction de distance, sous les contraintes définies dans l'équation 1.

Tirage après élimination de la base des échantillons précédemment tirés

La situation est la suivante : un échantillon d'établissements a été sélectionné pour participer à une évaluation ; un deuxième échantillon doit être tiré pour une autre évaluation. Nous souhaitons éviter que des établissements soient interrogés deux fois. Il s'agit donc de gérer le non-recouvrement entre les échantillons et d'assurer également un tirage du deuxième échantillon. Nous nous concentrons ici sur le non-recouvrement des échantillons mais notons qu'une approche plus générale incluant un taux de recouvrement non nul (pour permettre des analyses croisées entre enquêtes) dans un cadre de tirage équilibré est en cours de développement avec une application à des données issues d'évaluations standardisées (Christine & Rocher, 2012).

Notations

Un échantillon S_1 a été tiré. Il est connu et les probabilités d'inclusion des établissements π_j^1 sont également connues. On souhaite alors tirer un échantillon S_2 dans la population U avec les probabilités π_j^2 , mais sans aucun recouvrement avec l'échantillon S_1 . On va donc tirer l'échantillon S_2 dans la population $U(S_1)$, c'est-à-dire la population U privée des établissements de l'échantillon S_1 qui appartiennent à U . Notons d'emblée que S_1 n'a pas nécessairement été tiré dans U , mais potentiellement dans une autre population, plus large ou plus réduite ; cela n'affecte en rien la formulation envisagée ici. Notons également que l'indice j est utilisé ici : il concerne les établissements et non les élèves, représentés par l'indice i .

Il s'agit donc de procéder à un tirage conditionnel. On note π_j^{2/S_1} les probabilités d'inclusion conditionnelles des établissements dans le second échantillon S_2 , sachant que le premier échantillon est connu. Ces probabilités conditionnelles peuvent s'écrire :

$$\pi_j^{2/S_1} = \begin{cases} \lambda_j & \text{si } j \notin S_1 \\ 0 & \text{si } j \in S_1 \end{cases}, \text{ avec } \lambda_j \in [0, 1]$$

On a $\pi_j^2 = E(\pi_j^2/S_1) = \lambda_j(1 - \pi_j^1)$ d'où $\lambda_j = \frac{\pi_j^2}{1 - \pi_j^1}$

Condition fondamentale

Comme il s'agit d'une probabilité, la condition fondamentale est que $\lambda_j \in [0, 1]$. Comme $\lambda_j = \frac{\pi_j^2}{1 - \pi_j^1}$, la condition est en fait que

$$\pi_j^1 + \pi_j^2 \leq 1$$

Dans certains cas, par exemple des strates souvent sur-représentées comme les établissements situés dans des zones spécifiques concernant peu d'élèves (ex : REP+), cette condition pourrait ne pas être satisfaite. Cependant, de façon concrète, la condition a toujours été respectée dans les plans de sondage réalisés.

2.1.3 Calcul de précision : méthode

Les résultats des évaluations sont soumis à une variabilité qui dépend notamment des erreurs d'échantillonnage. Il est possible d'estimer statistiquement ces erreurs d'échantillonnage, appelées erreurs standard.

On note Y la variable d'intérêt (typiquement le score obtenu à une évaluation) et \hat{Y} l'estimateur de la moyenne de Y , qui constitue un estimateur essentiel sur lequel nous insistons dans la suite, bien que d'autres soient également au centre des analyses, comme ceux concernant la dispersion. La méthode retenue est cependant applicable à différents types d'estimateurs.

Nous souhaitons estimer la variance de cet estimateur, c'est-à-dire $V(\hat{Y})$. En absence de formule théorique pour calculer $V(\hat{Y})$, il existe plusieurs procédures permettant de l'estimer, c'est-à-dire de calculer $\hat{V}(\hat{Y})$, l'estimateur de la variance d'échantillonnage. Il peut s'agir de méthodes de linéarisation des formules (Taylor) ou bien de méthodes empiriques (méthodes de réplcation, jackknife, etc.). Ces méthodes sont bien décrites dans la littérature. Le lecteur est invité à consulter Tillé (2001) ou Ardilly (2006).

Cependant, lorsqu'un calage sur marges a été effectué, il faut en tenir compte pour le calcul de la précision. Dans ce cas, la variance de \hat{Y} est asymptotiquement équivalente à la variance des résidus de la régression de la variable d'intérêt sur les variables de calage.

En pratique, pour estimer la variance d'échantillonnage de \hat{Y} , tenant compte du calage effectué, il convient alors d'appliquer la procédure suivante :

1. On effectue la régression linéaire de la variable d'intérêt sur les variables de calage, en pondérant par les poids initiaux. Les résidus e_i de cette régression sont calculés.
2. Les valeurs $g_i e_i$ sont calculées, où g_i représente le rapport entre les poids CALMAR (w_i) et les poids initiaux (d_i) : $g_i = \frac{w_i}{d_i}$
3. La variance d'échantillonnage de \hat{Y} est alors obtenue en calculant la variance d'échantillonnage de $g_i e_i$.

2.2 Echantillonnage

Champ

Le champ des évaluations CEDRE à l'école est celui des élèves de CM2 scolarisés dans des écoles publiques et privées sous contrat de France métropolitaine. Pour des raisons de coût, les écoles ayant moins de 6 élèves de CM2 sont exclues du champ.

Sont donc exclus du champ :

- Les TOM.
- Les écoles hors contrat.
- Les écoles à l'étranger.
- Les écoles spécialisées.
- Les écoles de moins de 6 élèves de CM2.
- Les DOM.

La base de sondage est relativement pauvre en informations dans le premier degré. Nous disposons cependant d'informations sur les établissements scolaires, comme le secteur d'enseignement.

Stratification

La stratification prend en compte à la fois la taille et le secteur d'enseignement de l'école :

1. Écoles publiques hors ZEP (14 élèves ou plus)
2. Écoles publiques en ZEP (14 élèves ou plus)
3. Écoles privées (14 élèves ou plus)
4. Écoles publiques hors ZEP (entre 6 et 13 élèves)
5. Écoles publiques en ZEP (entre 6 et 13 élèves)
6. Écoles privées (entre 6 et 13 élèves)

Modalités de sélection

Le tirage est à deux degrés. Le premier degré de sondage est composé d'écoles tirées dans chaque strate avec allocation proportionnelle. La macro CUBE de l'INSEE est utilisée pour garantir des échantillons équilibrés sur la base de sondage selon certaines variables.

Dans chacune des strates, le tirage est équilibré sur la variable suivante :

- Le nombre total d'élèves de CM2

Le deuxième degré consiste à interroger tous les élèves de l'école sélectionnée (tirage par grappe).

Echantillon 2016

En pratique, on ne dispose pas de l'information sur la langue étrangère enseignée dans les écoles primaires. Dans les 10 académies concernées par l'enquête CEDRE LV CM2 - Allemand, on retire les écoles dans lesquelles les élèves apprennent l'allemand. Dans les autres académies, on considère que dans toutes les écoles les élèves apprennent l'anglais. L'échantillon vise 4 000 élèves répartis proportionnellement selon les trois strates.

Base de sondage

Le tableau 6 présente les exclusions dans la population ciblée.

Tableau 6 – Exclusions pour la base de sondage - CEDRE Anglais CM2 2016

	etab	eleves	etab exclus	eleves exclus
Ecoles accueillant des élèves de CM2	32 578	818 980		
On retire les TOM	32 387	814 737	191	4 243
On retire les écoles hors contrat	32 061	811 819	326	2 918
On retire les écoles spécialisées	32 041	811 274	20	545
On retire les DOM	31 111	773 119	930	38 155
On retire les petites écoles (< 6 CM2)	28 787	764 797	2 324	8 322
Total des exclusions CEDRE			3 791	54 183
Base CEDRE CM2	28 787	764 797		
Hors Ecoles qui font allemand (10 acad)	27 453	725 687	1 334	39 110
On retire PIRLS, B4, SOCLE CE2 tablettes et CEDRE CM2 All			1 069	37 711
Base de tirage CEDRE LV Anglais CM2	26 384	687 976		

Le tableau 7 présente la répartition de la population ciblée selon le secteur d'enseignement.

Tableau 7 – Répartition base de sondage - CEDRE Anglais CM2 2016

	strate	Ecoles	Elèves
1. Public hors EP		21 589	544 376
2. Public EP		3 040	104 501
3. Privé		4 158	115 920
Total		28 787	764 797

Échantillon

Le tableau 8 présente la répartition de l'échantillon selon le secteur d'enseignement. Au total, 153 écoles ont été sélectionnées.

Tableau 8 – Répartition dans l'échantillon - CEDRE Anglais CM2 2016

	strate	Ecoles	Eèves
1. Public hors EP		113	2 807
2. Public EP		17	588
3. Privé		23	675
Total		153	4 070

2.3 Etat des lieux de la non-réponse

2.3.1 Non-réponse totale

Parmi la non-réponse totale, nous distinguons selon la non-réponse d'écoles entières ou la non-réponse d'élèves dans les écoles participantes. Les chiffres suivants ont été observés pour 2016. Tout d'abord, 95,5 % des écoles de l'échantillon ont répondu à l'évaluation (tableau 9).

91,8 % des effectifs attendus en évaluation de l'écrit et 91,4 % des effectifs attendus en évaluation de l'oral ont participé à l'évaluation. Les élèves qui ont passé les deux mais dont on a pas pu rattacher les deux parties sont comptés en non réponse ici. Au final, 77,1 % des effectifs attendus ont participé (tableau 10).

Tableau 9 – Non-réponse des écoles - CEDRE Anglais CM2 2016

strate	Nb écoles attendues	Nb écoles répondantes	% de écoles répondantes
1- public hors EP	111	108	97,3%
2- EP	17	16	94,1%
3- privé	23	20	87,0%
Total	151	144	95,5%

Tableau 10 – Non-réponse globale - CEDRE Anglais CM2 2016

strate	Nb élèves attendus	Nb élèves répondants	% élèves répondants
1- public hors EP	2 743	2 229	81,3%
2- EP	588	418	71,1%
3- privé	675	440	65,2%
Total	4 006	3 087	77,1%

2.3.2 Valeurs manquantes et imputation

Dans le cas où certaines données sont manquantes, nous procédons à des imputations. Cela concerne uniquement les variables sexe et année de naissance, afin de pouvoir réaliser des statistiques selon ces variables sur l'échantillon complet, quelle que soit l'analyse. Nous imputons aléatoirement les valeurs manquantes de ces deux variables, de manière à respecter la répartition des répondants.

2.3.3 Non-réponse partielle et terminale

Lorsque des non-réponses sont observées aux items, nous distinguons les cas suivants :

- La non-réponse partielle : un élève n'a pas répondu à certains items dans le cahier.
- La non-réponse terminale : un élève s'est arrêté avant la fin du cahier soit par manque de temps soit par abandon.

Dans le premier cas, les non-réponses sont traitées comme des échecs (code "0"). Le second cas conduit à déterminer des règles. Nous considérons que si un élève a passé moins de 50 % d'une séquence, il n'a pas vu la séquence, les valeurs manquantes sont donc traitées de manière structurelle (code "s"). La non-réponse terminale a été étudiée par séquence et par cahier. Si un élève a passé moins de 50 % d'une séquence, on considère qu'il n'a pas vu la séquence (code "s").

2.3.3.a Compréhension de l'écrit et expression écrite

Les cahiers d'évaluation de l'écrit sont composés d'une séquence d'évaluation de la compréhension de l'écrit et d'une séquence d'évaluation de l'expression écrite. Parmi les élèves ayant de la non-réponse terminale, il y a en moyenne 4,9 pour la 1ère séquence et 4,6 pour la deuxième séquence.

Au final, on considère que :

- 111 élèves n'ont pas vu la séquence 1 dont :
 - 81 n'ont répondu à aucun item de la séquence
 - 30 ont répondu à moins de 50 % de la séquence
- 129 élèves n'ont pas vu la séquence 2 dont :
 - 87 n'ont répondu à aucun item de la séquence
 - 42 ont répondu à moins de 50 % de la séquence

2.3.3.b Compréhension de l'oral

Les cahiers d'évaluation de la compréhension de l'oral sont composés de deux séquences. Parmi les élèves ayant de la non-réponse terminale, il y a en moyenne 2,3 pour la 1ère séquence et 2,2 pour la deuxième séquence.

- 82 élèves n'ont pas vu la séquence 1 dont :
 - 80 n'ont répondu à aucun item de la séquence
 - 2 ont répondu à moins de 50 % de la séquence
- 87 élèves n'ont pas vu la séquence 4 dont :
 - 84 n'ont répondu à aucun item de la séquence
 - 3 ont répondu à moins de 50 % de la séquence

Les élèves dont les quatre séquences sont codées en "s" sont considérés comme de la non réponse totale. C'est le cas pour 11 élèves.

2.4 Redressement

Pour tenir compte de la non réponse, l'échantillon a été redressé à l'aide d'un calage sur marge. Préalablement au calage, on effectue tout d'abord une post-stratification. Puis, deux variables de calage sont utilisées :

- la répartition selon le sexe dans la population ;
- la répartition selon le retard scolaire.

Les élèves ayant passé les séquences d'évaluation de l'écrit ne sont pas forcément les mêmes que ceux ayant passé les séquences d'évaluation de l'oral. Un score à l'écrit et un score à l'oral sont calculés pour chaque élève, et on traitera séparément ces deux variables lors des analyses. Pour cette raison, deux calages sont effectués (11 et 12).

Tableau 11 – Comparaison entre les marges de l'échantillon et les marges dans la population - Ecrit

	Modalité ou variable	Marge échantillon	Marge population	Pourcentage échantillon	Pourcentage population
Retard	1	69 705.70	65 789	10.33	9.75
	2	604 807.34	608 724	89.67	90.25
Sexe	1	349 987.85	344 030	51.89	51.00
	2	324 525.19	330 483	48.11	49.00
Strate	1	471 012.04	471 012	69.83	69.83
	2	91 299.00	91 299	13.54	13.54
	3	112 202.00	112 202	16.63	16.63

Tableau 12 – Comparaison entre les marges de l'échantillon et les marges dans la population - Oral

	Modalité ou variable	Marge échantillon	Marge population	Pourcentage échantillon	Pourcentage population
Retard	1	69 776.52	65 789	10.34	9.75
	2	604 736.47	608 724	89.66	90.25
Sexe	1	349 401.32	344 030	51.80	51.00
	2	325 111.67	330 483	48.20	49.00
Strate	1	471 011.98	471 012	69.83	69.83
	2	91 299.00	91 299	13.54	13.54
	3	112 202.01	112 202	16.63	16.63

2.5 Précision

L'erreur standard (se) peut être calculée sur le score moyen de chaque année (tableau 13).

Tableau 13 – Scores moyens et erreurs standard associées - CEDRE Anglais CM2

	Année	Score moyen	Erreur standard
Compréhension de l'écrit	2004	250	1.73
	2010	272.4	1.73
	2016	277.3	1.86
Expression écrite	2004	250	1.87
	2010	265.1	1.43
	2016	259.5	1.87
Compréhension de l'oral	2004	250	2.17
	2010	260.7	2.16
	2016	257.5	1.92

Pour savoir par exemple si l'évolution entre 2010 et 2016 est significative, il faut calculer la valeur suivante :

$$\frac{|\hat{Y}_{2016} - \hat{Y}_{2010}|}{\sqrt{se_{\hat{Y}_{2016}}^2 + se_{\hat{Y}_{2010}}^2}} \quad (2)$$

Entre 2010 et 2016 en compréhension de l'écrit, on obtient une valeur de 1,93 (inférieure à 1,96). Cela signifie que la hausse du score moyen n'est pas statistiquement significative, tout comme en compréhension de l'oral (la valeur obtenue est 1.08). En expression écrite, la baisse du score moyen est significative (2.39).

En compréhension de l'écrit, de l'oral, et en expression écrite, les évolutions des scores entre 2004 et 2010, et entre 2004 et 2016 sont toutes significatives.

Les erreurs standards sont également calculées pour les répartitions dans les différents groupes de niveaux (tableaux 14 et 15).

Tableau 14 – Répartition en % dans les groupes de niveaux - CEDRE Anglais CM2

	Année	Groupe <1	Groupe 1	Groupe 2	Groupe 3	Groupe 4	Groupe 5
Compréhension de l'écrit	2004	4.4	10.6	26.3	31.1	17.6	10.0
	2010	1.7	6.5	18.2	28.9	25.4	19.4
	2016	0.3	3.5	17.9	32.2	24.3	21.8
Expression écrite	2004	4.3	13.0	27.2	25.3	21.8	8.4
	2010	4.1	8.4	17.0	27.1	25.9	17.5
	2016	3.4	6.7	20.7	35.6	22.5	11.2
Compréhension de l'oral	2004	2.5	12.5	32.4	30.5	12.0	10.1
	2010	4.7	12.1	21.2	24.0	18.6	19.4
	2016	2.0	11.0	26.1	28.3	17.9	14.7

Tableau 15 – Erreurs standards des répartitions en % dans les groupes de niveaux - CEDRE Anglais CM2

	Année	Groupe <1	Groupe 1	Groupe 2	Groupe 3	Groupe 4	Groupe 5
Compréhension de l'écrit	2004	0.3	0.5	0.8	0.7	0.6	1.0
	2010	0.2	0.5	0.8	0.7	0.7	1.1
	2016	0.1	0.4	1.0	1.0	1.0	1.4
Expression écrite	2004	0.4	0.6	0.9	0.6	0.8	1.0
	2010	0.3	0.5	0.6	0.7	0.7	0.9
	2016	0.4	0.6	1.2	1.1	1.1	1.0
Compréhension de l'oral	2004	0.3	0.7	1.0	0.8	0.7	1.1
	2010	0.4	0.7	0.9	0.7	0.8	1.3
	2016	0.3	1.0	1.1	0.9	1.0	1.3

Design effect

L'effet du plan de sondage (*Design Effect*) permet de rapporter l'erreur de mesure faite par un tirage spécifique à l'erreur de mesure qui aurait été faite en procédant à un sondage aléatoire simple (SAS) du même nombre d'élèves. Pour la moyenne d'une variable Y et un plan de sondage complexe P :

$$D_{eff} = \frac{V_P(\hat{Y})}{V_{SAS}(\hat{Y})} \quad (3)$$

Dans le cas d'un sondage en grappes, la précision est dégradée en comparaison d'un sondage aléatoire simple. Cela signifie qu'en 2016, en compréhension de l'écrit par exemple (tableau 16), un sondage aléatoire simple avec un effectif 2,02 fois moins important aurait conduit au même niveau de précision.

Tableau 16 – Effet du plan de sondage - CEDRE Anglais CM2 2016- Compréhension de l'écrit

Année	Erreur Standard	Erreur SAS	<i>Design Effect</i>
2004	1.73	0.65	2.67
2010	1.73	0.76	2.27
2016	1.86	0.92	2.02

3 Analyse des items

3.1 Méthodologie

Pour une description générale de la méthodologie psychométrique employée dans les évaluations standardisées de compétences des élèves, le lecteur est invité à consulter Rocher (2015).

3.1.1 Approche classique

Dans un premier temps, nous posons quelques notations et nous présentons les principales statistiques descriptives utilisées pour décrire un test, issues de la « théorie classique des tests » que nous évoquons rapidement.

Réussite et score

On note n le nombre d'élèves ayant passé une évaluation composée de J items. On note Y_i^j la réponse de l'élève i ($i = 1, \dots, n$) à l'item j ($j = 1, \dots, J$). Dans notre cas, les items sont dichotomiques, c'est-à-dire qu'ils ne prennent que deux modalités (la réussite ou l'échec) :

$$Y_i^j = \begin{cases} 1 & \text{si l'élève } i \text{ réussit l'item } j \\ 0 & \text{si l'élève } i \text{ échoue à l'item } j \end{cases} \quad (4)$$

Le taux de réussite à l'item j est la proportion d'élèves ayant réussi l'item j . Il est noté p_j :

$$p_j = \frac{1}{n} \sum_{i=1}^n Y_i^j \quad (5)$$

Le taux de réussite d'un item renvoie à son niveau de difficulté. C'est certainement la caractéristique la plus importante, qui permet de construire un test de niveau adapté à l'objectif de l'évaluation, en s'assurant que les différents niveaux de difficulté sont balayés.

Le score observé à l'évaluation pour l'élève i , noté S_i , correspond au nombre d'items réussis par l'individu i :

$$S_i = \sum_{j=1}^J Y_i^j \quad (6)$$

La théorie classique des tests a précisément pour objet d'étude le score S_i obtenu par un élève à un test. Elle postule notamment que ce score observé résulte de la somme d'un score « vrai » inobservé et d'une erreur de mesure. Un certain

nombre d'hypothèses portent alors sur le terme d'erreur (pour plus d'informations, cf. par exemple Laveault et Gregoire, 2002).

Fidélité

Dans le cadre de la théorie classique des tests, la fidélité (*reliability*) est définie comme la corrélation entre le score observé et le score vrai : le test est fidèle, lorsque l'erreur de mesure est réduite. Une manière d'estimer cette erreur de mesure consiste par exemple à calculer les corrélations entre les différents sous-scores possibles : plus ces corrélations sont élevées, plus le test est dit fidèle¹.

Le coefficient α de Cronbach est un indice destiné à mesurer la fidélité de l'épreuve. Il est compris entre 0 et 1. Sa version « standardisée » s'écrit :

$$\alpha = \frac{J\bar{r}}{1 + (J - 1)\bar{r}} \quad (7)$$

où \bar{r} est la moyenne des corrélations inter-items.

De ce point de vue, cet indicateur renseigne sur la consistance interne du test. En pratique, une valeur supérieure à 0,8 témoigne d'une bonne fidélité².

Indices de discrimination

Des indices importants concernent le pouvoir discriminant des items. Nous présentons ici l'indice « r-bis point » ou coefficient point-bisérial qui est le coefficient de corrélation linéaire entre la variable indicatrice de réussite à l'item Y^j et le score S .

Appelé également « corrélation item-test », il indique dans quelle mesure l'item s'inscrit dans la dimension générale. Une autre manière de l'envisager consiste à le formuler en fonction de la différence de performance constatée entre les élèves qui réussissent l'item et ceux qui l'échouent.

1. Notons au passage que la naissance des analyses factorielles est en lien avec ce sujet : Charles Spearman cherchait précisément à dégager un facteur général à partir de l'analyse des corrélations entre des scores obtenus à différents tests.

2. La littérature indique plutôt un seuil de 0,70 (Peterson, 1994). Cependant, comme le montre la formule ci-dessus, le coefficient α est lié au nombre d'items, qui est important dans les évaluations conduites par la DEPP afin de couvrir les nombreux éléments des programmes scolaires. Des facteurs de correction existent néanmoins et permettent de comparer des tests de longueur différentes.

En effet, on peut montrer que

$$r_{bis-point}(j) = corr(Y^j, S) = \frac{\bar{S}_{(j1)} - \bar{S}_{(j0)}}{\sigma_S} \sqrt{p_j(1 - p_j)} \quad (8)$$

où $\bar{S}_{(j1)}$ est le score moyen sur l'ensemble de l'évaluation des élèves ayant réussi l'item j , $\bar{S}_{(j0)}$ celui des élèves l'ayant échoué et σ_S est l'écart-type des scores.

C'est donc bien un indice de discrimination, entre les élèves qui réussissent et ceux qui échouent à l'item. En pratique, on préfère s'appuyer sur les $r_{bis-point}$ corrigés, c'est à dire calculés par rapport au score à l'évaluation privée de l'item considéré. Une valeur inférieure à 0,2 indique un item peu discriminant (Laveault et Grégoire, 2002).

3.1.2 Analyse factorielle des items

L'analyse factorielle permet d'étudier la structure des données et, plus particulièrement, la structure des corrélations entre les variables observées (ou manifestes)³. Il s'agit d'identifier les différentes dimensions sous-jacentes aux réussites observées et surtout d'évaluer le poids de la dimension principale, dans la mesure où c'est une optique unidimensionnelle qui sera envisagée lors de la modélisation.

Dans le cas où les items sont dichotomiques, la matrice des corrélations entre items est en fait la matrice des coefficients ϕ , qui sont bornés selon les taux de réussite aux items (Rocher, 1999). Une analyse factorielle basée sur cette matrice peut donc montrer quelques faiblesses : des facteurs « artefactuels » sont susceptibles d'apparaître, en lien avec le niveau de difficulté des items et non avec les dimensions auxquelles ils se rapportent. De plus, d'un point de vue théorique, certaines hypothèses utiles pour l'estimation, comme la normalité des variables, ne sont pas envisageables.

L'optique retenue est alors de se ramener à un modèle linéaire : les variables observées catégorielles sont considérées comme la manifestation de variables latentes continues.

3. Notons qu'il s'agit ici d'analyse factorielle en facteurs communs et spécifiques et non d'analyse factorielle géométrique de type ACP ou ACM (pour des détails, consulter Rocher, 2013)

Les réponses à un item dichotomique sont définies de la manière suivante :

$$y_{ij} = \begin{cases} 0 & \text{si } z_{ij} \leq \tau_j \\ 1 & \text{si } z_{ij} > \tau_j \end{cases} \quad (9)$$

La réponse y_{ij} de l'élève i à l'item j est incorrecte tant que la variable latente Z_j reste en deçà d'un certain seuil τ_j , qui dépend de l'item. Au-delà de ce seuil, la réponse est correcte.

L'analyse factorielle des items consiste donc en une analyse factorielle linéaire sur les variables continues Z_j . Deux modèles sont donc considérés. D'une part, une variable latente continue et conditionnant la réponse à l'item est fonction linéaire de facteurs communs et d'un facteur spécifique. D'autre part, un modèle de seuil représente la relation non linéaire entre la variable latente et la réponse à l'item. Ce procédé permet de se ramener à une analyse factorielle linéaire, à la différence que les variables Z_j ne sont pas connues. Il s'agit donc d'estimer la matrice de corrélation de ces variables, sous certaines hypothèses.

Considérons le lien entre deux items j et k . Si les variables latentes correspondantes Z^j et Z^k sont distribuées selon une loi normale bivariée, il est possible d'estimer le coefficient de corrélation linéaire de ces deux variables à partir du tableau croisant les deux items. C'est le coefficient de corrélation tétrachorique – ou polychorique dans le cas d'items polytomiques. L'estimation de ce coefficient par le maximum de vraisemblance requiert la résolution d'une double intégrale (pour les détails de l'estimation pour deux items dichotomiques, cf. Rocher, 1999). Pour plus de deux items, il devient difficile d'estimer de la même manière les coefficients de corrélation à partir de la distribution conjointe des items qui est une loi normale multivariée. C'est pourquoi les coefficients de corrélation tétrachorique sont estimés séparément pour chaque couple d'items. Ce procédé a le désavantage de conduire à une matrice de covariances qui n'est pas nécessairement semi-définie positive, donc potentiellement non inversible.

3.2 Codage des réponses aux items

3.2.1 Valeurs manquantes

Trois types de valeurs manquantes sont distinguées :

- Valeurs manquantes structurelles : l'élève n'a pas vu l'item. C'est le cas pour les cahiers tournants, où les élèves ne voient pas tous les items. Dans ce cas, on considère l'item comme *non administré*, l'absence de réponse n'est alors pas considérée comme une erreur.
- Absence de réponse : l'élève a vu l'item mais n'y a pas répondu. L'absence de réponse est alors considérée comme une erreur de la part de l'élève.

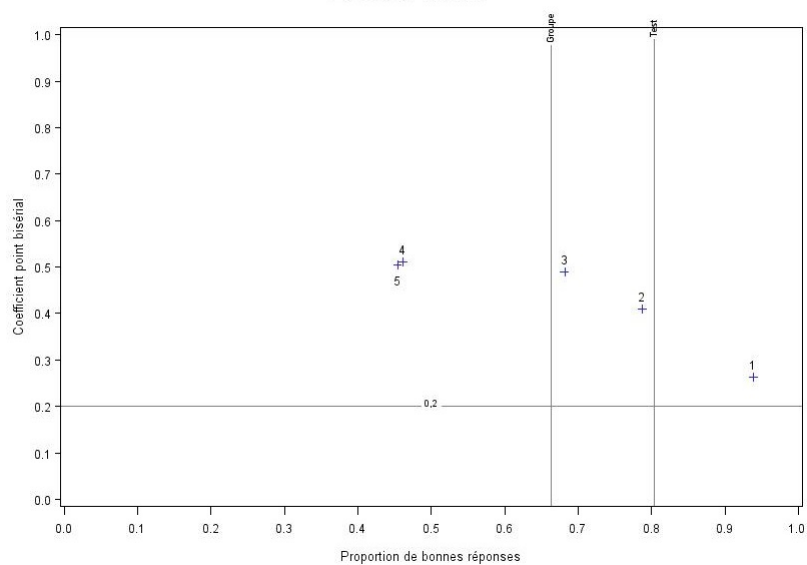
- Non-réponse terminale : l'élève s'est arrêté au cours de l'épreuve, potentiellement en raison d'un manque de temps. Des choix sont effectués pour déterminer le traitement de ces valeurs. Nous considérons que si un élève a passé moins de 50 % d'une séquence, il n'a pas vu la séquence, les valeurs manquantes sont alors traitées de manière structurelle. Sinon, elles sont traitées comme des échecs.

3.2.2 Regroupement des items

Les séries d'items comportant seulement deux réponses, comme les Vrai/Faux, font l'objet d'un traitement spécifique (cf. l'exemple 2 donné au paragraphe ??). Les items de ce type sont regroupés pour former un seul item à réponse binaire (réussite ou échec). En effet, la plus forte potentialité de réponse au hasard et l'inter-dépendance des items fragilisent leur utilisation individuelle.

Le regroupement de ces items consiste à faire la somme des indicatrices de réussite et à déterminer un seuil de maîtrise. Une visualisation graphique est utilisée pour fixer les scores « seuils » (cf. figure 1). Ce graphique représente le taux de réussite pour chaque seuil possible en fonction de la discrimination obtenu pour le seuil. Il permet de choisir la combinaison la mieux adaptée. Le score seuil doit préserver la discrimination de l'item regroupé et la difficulté peut être modulée en fonction des objectifs.

Figure 1 – Représentation graphique utilisée pour le regroupement d'items



Note de lecture : L'item présenté ici est une série de cinq questions de type « Vrai/Faux ». Chaque croix représente l'item correspondant au seuil de réussite retenu. Par exemple, si la réussite à l'ensemble est attribuée dès lors qu'une seule question est réussie, l'item obtenu a un taux de réussite d'environ 95 % et un coefficient biserial d'environ 0,26. Si le seuil de réussite est fixé à 3 questions réussies sur 5, alors le taux de réussite baisse mécaniquement (autour de 65 % qui est le taux de réussite obtenu à l'ensemble des questions de cet item).

3.2.3 Traitement des données et correction des questions ouvertes

Tous les cahiers recueillis dans le cadre de cette opération ont été scannés par une société extérieure. Les réponses aux questions à choix multiples ainsi que les grilles d'évaluation remplies par les professeurs lors des séquences de travaux pratiques ont été numérisées et les codes de réponses stockés dans un fichier. En ce qui concerne les questions ouvertes, demandant une rédaction plus ou moins longue de la part des élèves (explication, schématisation...), elles ont été découpées en « imagettes » puis transmises au ministère afin d'être intégrées dans un logiciel de correction à distance (cf. encadré « AGATE »). Celui-ci nécessite la formation technique des correcteurs et l'élaboration d'un cahier des charges strict de corrections pour limiter la subjectivité des corrections. Une fois la correction terminée, les codes saisis par les correcteurs ont été stockés dans un fichier puis associés à ceux issus des réponses aux QCM.

AGATE : un outil de correction à distance des questions ouvertes

Objectifs

Le logiciel AGATE, qui a été développé par les informaticiens de la DEPP, permet une correction à distance des questions ouvertes. Le principe général du logiciel est de soumettre un lot d'imagettes (image scannée de la réponse d'un élève) à un groupe de correcteurs tout en paramétrant des contraintes de double correction et/ou d'auto-correction. Lorsque deux correcteurs corrigent la même imagette, il arrive parfois qu'il y ait une différence de codage. Cette imagette est alors proposée au superviseur qui arbitre et valide l'un des deux codages. Ce jeu de codages multiples incrémente des compteurs (temps de connexion, avancement général et taux d'erreur) qui sont autant d'indicateurs pour suivre la correction. A noter qu'un processus de déconnexion automatique d'un correcteur existe si le superviseur se rend compte d'un trop grand nombre d'erreurs de correction. Ce logiciel est utilisé depuis 2004 par le bureau des évaluations de la DEPP. Il a permis d'intégrer des questions ouvertes dans des évaluations à grandes échelles, aussi bien aux évaluations nationales qu'aux évaluations internationales telles PISA, TIMSS ou PIRLS. Les correcteurs n'ont plus à manipuler un nombre très important de cahiers et peuvent travailler de manière autonome lorsqu'ils le souhaitent, tout en maintenant un contact entre eux et les responsables de l'évaluation afin d'assurer une meilleure fiabilité de la correction.

Principes fonctionnels

Le chef de projet paramètre la session de correction. Il définit les groupes de correcteurs et supervise chaque groupe. Il intègre et vérifie les items mis en correction et ajuste les paramètres de double correction. Son rôle consiste également à répondre aux questions des correcteurs par le biais d'une messagerie intégrée au logiciel et à communiquer sa réponse également aux autres correcteurs. Le superviseur gère son groupe de correcteurs. Il anime la session de formation, qui consiste d'une part à communiquer aux télécorrecteurs une grille de correction très précises et d'autre part à corriger collectivement à blanc un nombre défini d'imagettes pour s'assurer de la compréhension et de la bonne mise en oeuvre des consignes. Puis, pendant la télécorrection, il arbitre les litiges lors des doubles-corrrections. Le correcteur corrige les items en portant un codage de réussite/erreur sur chaque item. En cas de doute, il peut se référer à son superviseur de groupe. Une messagerie interne complète le dispositif et permet un échange de point de vue entre les différents acteurs.

3.3 Résultats

3.3.1 Pouvoir discriminant des items

- En compréhension de l'écrit, le calcul des indices de discrimination nous a amené à éliminer 12 items dont les indices *rbis-point* étaient trop faibles : 2 items de 2004, 9 items de 2016, et 1 item communs aux trois années.
- En compréhension de l'oral, 45 items ont été éliminés : 35 items de 2004, 1 item de 2010, 7 items de 2016, et 2 items communs à 2010 et 2016.
- En expression écrite, 4 items ont été supprimés : 1 item de 2010, 1 item commun à 2004 et 2010, 1 item commun à 2010 et 2016, et 1 item commun aux trois années.

4 Modélisation

4.1 Méthodologie

4.1.1 Modèle de réponse à l'item

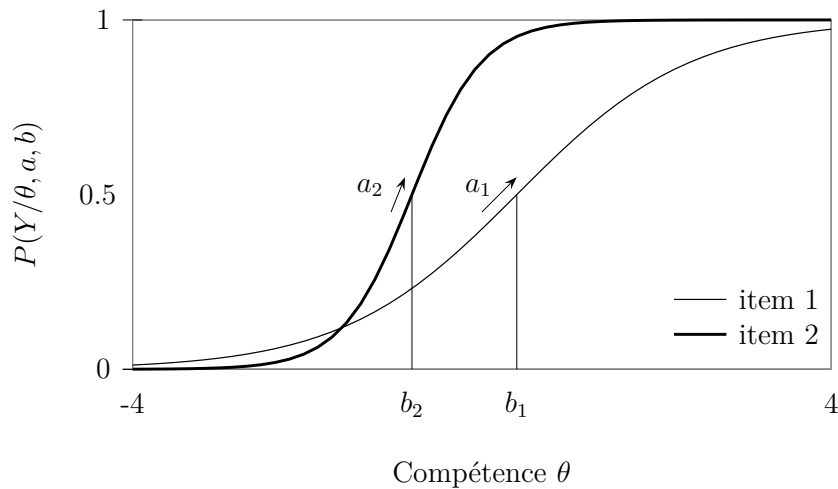
Le modèle de mesure utilisé est un modèle de réponse à l'item à deux paramètres avec une fonction de lien logistique (MRI 2PL) :

$$P_{ij} = P(Y_i^j = 1 | \theta_i, a_j, b_j) = \frac{e^{1,7a_j(\theta_i - b_j)}}{1 + e^{1,7a_j(\theta_i - b_j)}} \quad (10)$$

où la probabilité P_{ij} que l'élève i réussisse l'item j est fonction du niveau de compétence θ_i de l'élève i , du niveau de difficulté b_j de l'item j , ainsi que de la discrimination de l'item a_j ($a_j > 0$). La constante 1,7 est introduite pour rapprocher la fonction sigmoïde de la fonction de répartition de la loi normale.

La figure 2 représente les courbes caractéristiques de deux items selon cette modélisation.

Figure 2 – Modèle de réponse à l'item - 2 paramètres



Note de lecture : la probabilité de réussir l'item (en ordonnées) dépend du niveau de compétence (en abscisse). L'item 1 en trait fin est plus difficile que l'item 2 en trait plein ($b_1 > b_2$), et il est moins discriminant ($a_1 < a_2$).

L'avantage de ce type de modélisation, c'est de séparer deux concepts-clé, à savoir la difficulté de l'item et le niveau de compétence de l'élève. Les MRI ont un intérêt pratique pour la construction de tests et la comparaison entre différents groupes d'élèves : si le modèle est bien spécifié sur un échantillon donné, les paramètres des items – en particulier leurs difficultés – peuvent être considérés comme fixes et applicables à d'autres échantillons dont il sera alors possible de déduire les paramètres relatifs aux élèves – en particulier, leur niveau de compétence. Pour une présentation générale, le lecteur est invité à consulter Rocher (2015).

Autre avantage : le niveau de compétence des élèves et la difficulté des items sont placés sur la même échelle, par le simple fait de la soustraction ($\theta_i - b_j$). Cette propriété permet d'interpréter le niveau de difficulté des items par rapprochement avec le continuum de compétence. Ainsi, les élèves situés à un niveau de compétence égal à b_j auront 50 % de chances de réussir l'item, ce que traduit visuellement la représentation des courbes caractéristiques des items (CCI) selon ce modèle (figure 2).

4.1.2 Procédures d'estimation

L'estimation est conduite en deux temps : l'estimation des paramètres des items puis l'estimation des θ en considérant les paramètres des items comme fixes. Nous donnons ici des éléments concernant ces procédures.

Estimation des paramètres des items

Nous reprenons les notations de l'équation (10) qui formule la probabilité P_{ij} d'un élève i de répondre correctement à un item j dans le cadre d'un modèle de réponse à l'item, avec les items sont dichotomiques.

Notons tout d'abord que les modèles présentés ne sont pas identifiables. En effet, les transformations $\theta_i^* = A\theta_i + B$, $b_j^* = Ab_j + B$ et $a_j^* = a_j/A$ avec A et B deux constantes ($A > 0$), conduisent aux mêmes valeurs des probabilités. Dans CEDRE, nous levons l'indétermination en standardisant la distribution des θ pour les données du premier cycle (en l'occurrence, moyenne de 250 et écart-type de 50 pour l'année 2004).

Sous l'hypothèse d'indépendance locale des items⁴, la fonction de vraisemblance s'écrit :

$$L(\mathbf{y}, \xi, \theta) = \prod_{i=1}^n \prod_{j=1}^J P_{ij}^{y_{ij}} [1 - P_{ij}]^{1-y_{ij}} \quad (11)$$

4. Cette hypothèse signifie que les indicatrices de réussite des items sont indépendantes, conditionnellement au niveau de compétence θ . A niveau de compétence égal, deux items donnés ne sont pas corrélés : seule la compétence θ explique la corrélation entre deux items. Cette hypothèse est ainsi liée à l'hypothèse d'unidimensionnalité de θ (cf, Rocher, 2013).

où \mathbf{y} est le vecteur des réponses aux items (*pattern*), ξ est le vecteur des paramètres des items.

La procédure MML (*Marginal Maximum Likelihood*) est utilisée. Elle consiste à estimer les paramètres des items en supposant que les paramètres des individus sont issus d'une distribution fixée *a priori* (le plus souvent normale). La maximisation de vraisemblance est *marginale* dans le sens où les paramètres concernant les individus n'apparaissent plus dans la formule de vraisemblance.

Si θ est considérée comme une variable aléatoire de distribution connue, la probabilité inconditionnelle d'observer un *pattern* \mathbf{y}_i donné peut s'écrire :

$$P(\mathbf{y} = \mathbf{y}_i) = \int_{-\infty}^{+\infty} P(\mathbf{y} = \mathbf{y}_i | \theta_i) g(\theta_i) d\theta_i \quad (12)$$

avec g la densité de θ .

L'objectif est alors de maximiser la fonction de vraisemblance :

$$L = \prod_{i=1}^n P(\mathbf{y} = \mathbf{y}_i) \quad (13)$$

Cependant, l'annulation des dérivées de L par rapport aux a_j et aux b_j conduit à résoudre un système d'équations relativement complexe et à procéder à des calculs d'intégrales qui peuvent s'avérer très coûteux en termes de temps de calcul.

La résolution de ces équations est classiquement réalisée grâce à l'algorithme EM (*Expectation-Maximization*) impliquant des approximations d'intégrales par points de quadrature. L'algorithme EM est théoriquement adapté dans le cas de valeurs manquantes. Le principe général est de calculer l'espérance conditionnelle de la vraisemblance des données complètes (incluant les valeurs manquantes) avec les valeurs des paramètres estimées à l'étape précédente, puis de maximiser cette espérance conditionnelle pour trouver les nouvelles valeurs des paramètres. Le calcul de l'espérance conditionnelle nécessite cependant de connaître (ou de supposer) la loi jointe des données complètes. Une version modifiée de l'algorithme considère dans notre cas le paramètre θ lui-même comme une donnée manquante. Pour plus de détails, le lecteur est invité à consulter Rocher (2013).

En outre, ce cadre d'estimation permet aisément de traiter des valeurs manquantes structurelles, par exemple dans le cas de cahiers tournants ou bien dans le cas de reprise partielle d'une évaluation.

Estimation des niveaux de compétence

Une fois les paramètres des items estimés, ils sont considérés comme fixes et il est possible d'estimer les θ_i , par exemple *via* la maximisation de la vraisemblance donnée par l'équation (11).

Cependant, l'estimateur du maximum de vraisemblance, noté $\theta_i^{(ML)}$, est biaisé : les propriétés classiques de l'estimateur selon la méthode du maximum de vraisemblance ne sont pas vérifiées puisque le nombre de paramètres augmente avec le nombre d'observations. Ce biais vaut :

$$B(\theta_i^{(ML)}) = \frac{-J}{2I^2} \quad (14)$$

avec

$$I = \sum_{j=1}^J \frac{P'_{ij}{}^2}{P_{ij}(1-P_{ij})} = \sum_{j=1}^J a_j^2 P_{ij}(1-P_{ij})$$

et

$$J = \sum_{j=1}^J \frac{P'_{ij} P''_{ij}}{P_{ij}(1-P_{ij})} = \sum_{j=1}^J a_j^3 P_{ij}(1-P_{ij})$$

Pour obtenir un estimateur non biaisé, Warm (1989) a proposé de maximiser une vraisemblance pondérée $w(\theta)L(\mathbf{y}, \mathbf{a}, \mathbf{b}, \theta)$, en choisissant $w(\theta)$ de manière à ce que l'annulation de la dérivée du logarithme de la vraisemblance pondérée revienne à résoudre l'équation suivante :

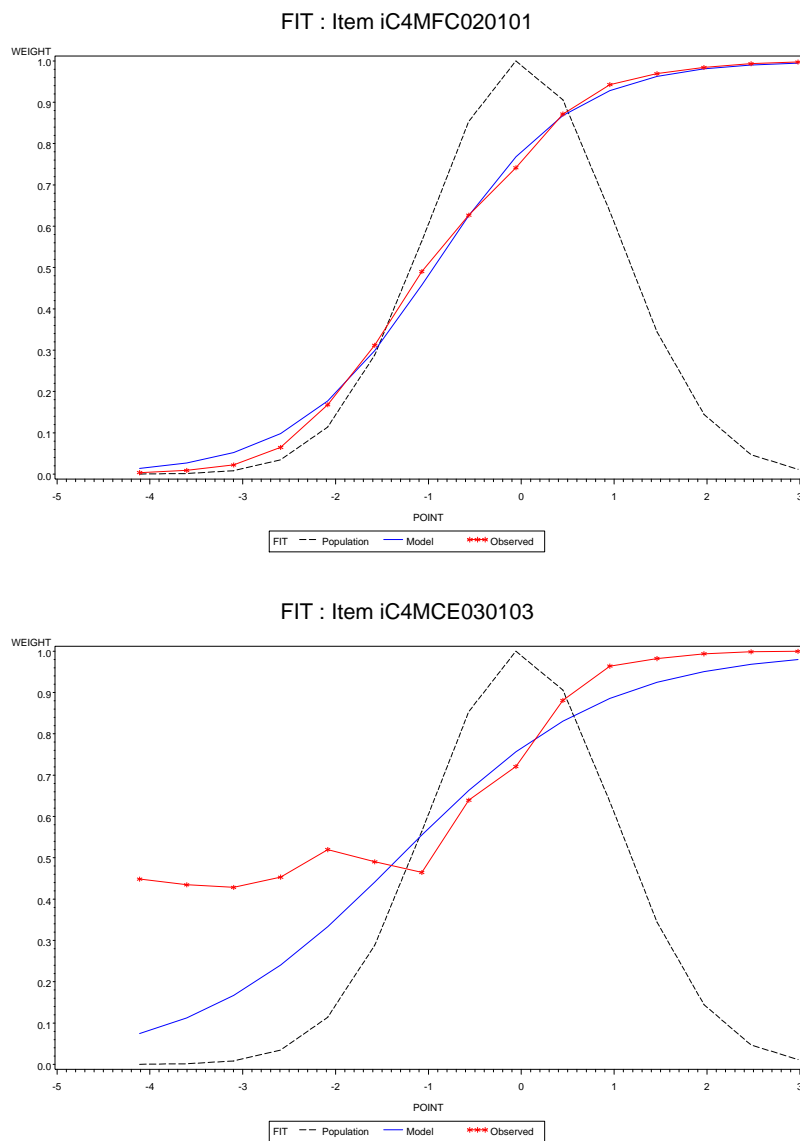
$$\frac{\partial \ln L}{\partial \theta_i} + \frac{J}{2I} = 0 \quad (15)$$

4.1.3 Indice d'ajustement (FIT)

L'ajustement des items au modèle est étudié. Graphiquement, cela revient à comparer les courbes caractéristiques estimées avec les résultats observés (cf. figure 3). Certaines procédures proposent de comparer directement les probabilités théorique avec les proportions de réussite de groupes d'élèves. Plus généralement, nous pouvons écrire les résidus de la manière suivante :

$$z_{ij} = \frac{Y_i^j - P_{ij}}{\sqrt{P_{ij}(1-P_{ij})}} \quad (16)$$

Figure 3 – Exemples d’ajustements (FIT)



Note de lecture : La courbe bleue représente la courbe caractéristique de l’item telle qu’estimée par le modèle. La courbe en rouge relie des points qui correspondent aux taux de réussite observé à cet item pour 15 groupes d’élèves de niveaux de compétence croissants. Enfin, la courbe en pointillée représente la distribution des niveaux de compétence.

Clairement, l’ajustement du modèle est excellent pour l’item présenté en haut. Il est très mauvais pour celui du bas.

Les carrés des résidus suivent typiquement une loi du χ^2 . L'indice *Infit* d'un item correspond à la moyenne pondérée des carrés des résidus, qui peut s'écrire :

$$Infit_j = \frac{1}{\sum_{i=1}^n w_{ij}} \sum_{i=1}^n w_{ij} z_{ij}^2 = \frac{1}{\sum_{i=1}^n w_{ij}} \sum_{i=1}^n (Y_i^j - P_{ij})^2 \quad (17)$$

avec le poids $w_{ij} = P_{ij}(1 - P_{ij})$. Une transformation de cet indice est utilisé de manière à obtenir une statistique suivant approximativement et empiriquement (le lien théorique n'est pas établi) une loi normale (Smith, Schumaker, & Bush, 1998).

4.1.4 Fonctionnement Différentiel d'Item (FDI)

Un fonctionnement différentiel d'item (FDI) apparaît entre des groupes d'individus dès lors qu'à niveau égal sur la variable latente mesurée, la probabilité de réussir un item donné n'est pas la même selon le groupe considéré. La question des FDI est importante car elle renvoie à la notion d'équité entre les groupes : un test ne doit pas risquer de favoriser un groupe par rapport à un autre.

Une définition formelle du FDI peut s'envisager à travers la propriété d'invariance conditionnelle : à niveau égal sur la compétence visée, la probabilité de réussir un item donné est la même quel que soit le groupe de sujets considéré. Formellement, un fonctionnement différentiel se traduit donc par :

$$P(Y | Z, G) \neq P(Y | Z) \quad (18)$$

où Y est le résultat d'une mesure de la compétence visée, typiquement la réponse à un item ; Z est un indicateur du niveau de compétence des sujets ; G est un indicateur de groupes de sujets.

Si la probabilité de réussite, conditionnellement au niveau mesuré, est différente selon les groupes d'élèves, alors il existe un fonctionnement différentiel.

En pratique, de très nombreuses méthodes ont été proposées afin d'identifier les FDI. Ces méthodes ont chacune des avantages en matière d'investigation des différents éléments pouvant conduire à l'apparition de ces FDI (Rocher, 2013). Dans le cas des évaluations standardisées menées à la DEPP, il s'agit avant tout d'identifier les fonctionnements différentiels pouvant apparaître entre deux moments de mesure, s'agissant des items repris à l'identique. Dans ce cas, les différentes méthodes d'identification donnent des résultats relativement proches.

Une stratégie très simple, employée dans CEDRE, consiste donc à comparer les paramètres de difficulté des items repris, estimés de façon séparée pour les deux

années. Si la difficulté d'un item a évolué, comparativement aux autres items, c'est le signe d'un fonctionnement différentiel, qui peut être lié par exemple à un changement de programmes ou de pratiques. Plus précisément, les paramètres des items sont estimés séparément pour les deux années, puis ajustés en tenant compte de la différence moyenne entre les deux séries de paramètres. La règle retenue pour identifier un FDI est celle d'un écart de paramètres de difficulté β d'au moins 0,5 (cf. Rocher, 2013 pour plus de détails).

4.1.5 L'information du test

Dans le cadre d'un modèle de réponse à l'item à deux paramètres, l'information d'un item j est définie par :

$$I_j(\theta) = (1,7a_j)^2 P_j(\theta)(1 - P_j(\theta)) \quad (19)$$

avec $P_j(\theta)$, la probabilité de réussite à l'item pour individu de compétence θ .

L'information moyenne du test pour un élève de compétence θ est la somme de l'information apporté par chaque item pour θ . La courbe d'information du test est tracée pour un ensemble de valeurs de θ . L'erreur de mesure étant inversement proportionnelle à l'information, cette courbe d'information permet de visualiser la précision avec laquelle le niveau de compétence θ des élèves est estimé.

4.2 Résultats

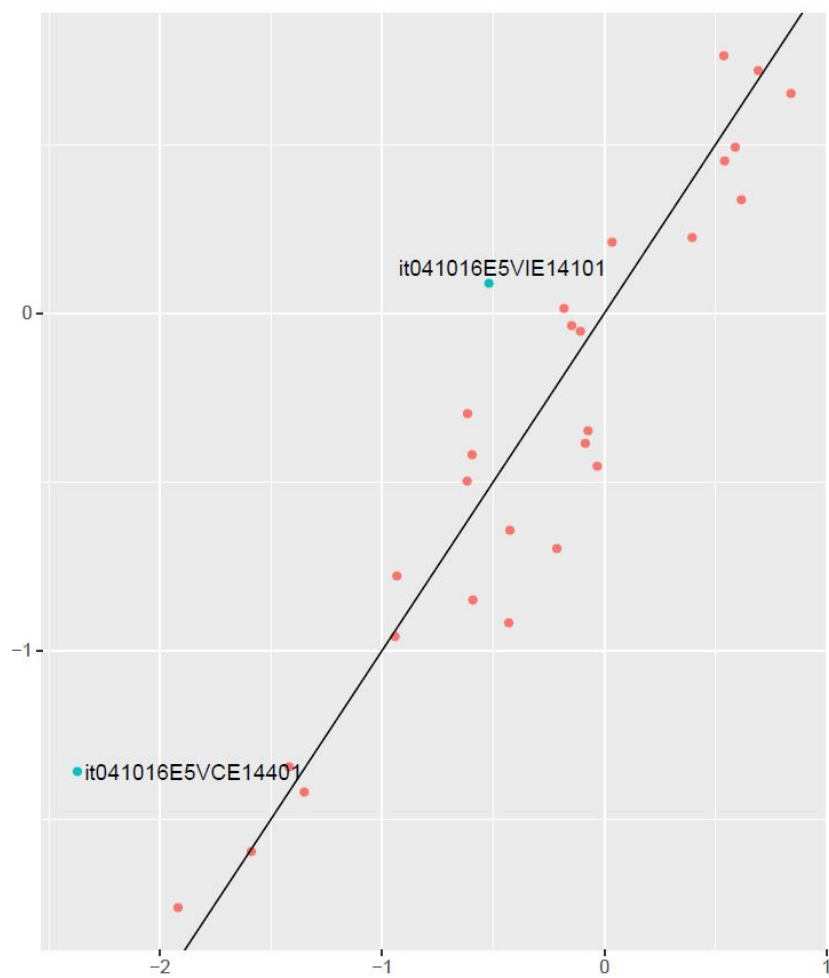
4.2.1 Identification des fonctionnements différentiels d'items (FDI)

L'analyse des FDI 2004-2010 a permis de détecter 2 items en compréhension de l'écrit (figure 4), et 2 items en expression écrite.

L'analyse des FDI 2010-2016 a permis de détecter 1 item en compréhension de l'écrit (figure 5), 7 items en compréhension de l'oral, et 3 items en expression écrite.

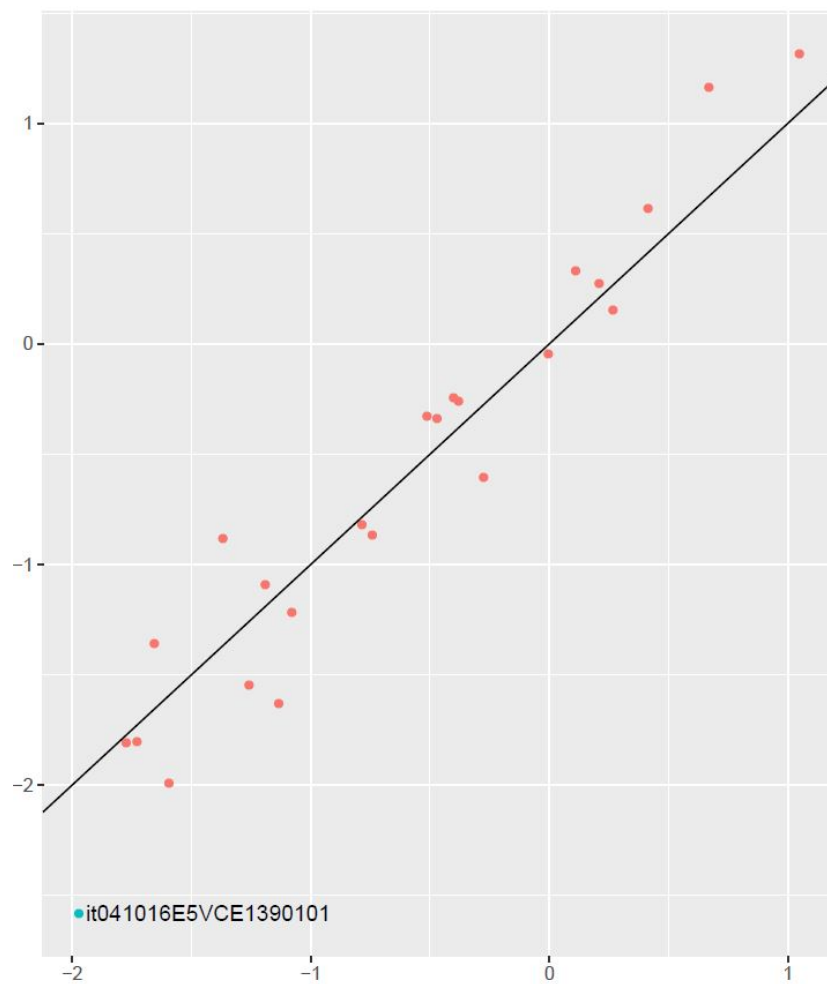
Ces items ont été éliminés des calculs.

Figure 4 – Comparaison des paramètres de difficulté 2004-2010 (CEDRE Anglais CM2)



Note de lecture : Les points sont les items. En abscisse figure la valeur des paramètres de difficulté estimés en 2004, et en ordonnée la la valeur des paramètres de difficulté estimés et ajustés pour l'année 2010. Les items présentant un FDI apparaissent en bleu.

Figure 5 – Comparaison des paramètres de difficulté 2010-2016 (CEDRE Anglais CM2)



Note de lecture : Les points sont les items. En abscisse figure la valeur des paramètres de difficulté estimés en 2010, et en ordonnée la la valeur des paramètres de difficulté estimés et ajustés pour l'année 2016. Les items présentant un FDI apparaissent en bleu.

4.2.2 Identification des items présentant un mauvais ajustement (FIT)

Aucun item présentant un mauvais ajustement n'a été détecté.

4.2.3 Bilan de l'analyse des items

4.2.3.a Compréhension de l'écrit

En considérant l'ensemble des items sur les trois années, il y avait au départ :

- 19 items communs aux trois années
- 11 items communs à 2004 et 2010
- 6 items communs à 2010 et 2016
- 17 items de 2004
- 2 items de 2010
- 66 items de 2016

Cela représente 121 items passés par les élèves en tout, dont 91 en 2016.

Après suppression des items présentant un mauvais Rbis, un fonctionnement différentiel ou un mauvais ajustement, il reste :

- 18 items communs aux trois années
- 9 items communs à 2004 et 2010
- 5 items communs à 2010 et 2016
- 15 items de 2004
- 2 items de 2010
- 57 items de 2016

109 items sont donc conservés dans l'analyse, dont 80 utilisés dans l'évaluation 2016.

4.2.3.b Compréhension de l'oral

En considérant l'ensemble des items sur les trois années, il y avait au départ :

- 7 items communs aux trois années
- 2 items communs à 2004 et 2010
- 18 items communs à 2010 et 2016
- 158 items de 2004
- 11 items de 2010
- 65 items de 2016

Cela représente 261 items passés par les élèves en tout, dont 90 en 2016.

Après suppression des items présentant un mauvais Rbis, un fonctionnement différentiel ou un mauvais ajustement, il reste :

- 7 items communs aux trois années
- 2 items communs à 2004 et 2010
- 9 items communs à 2010 et 2016
- 123 items de 2004
- 10 items de 2010
- 58 items de 2016

216 items sont donc conservés dans l'analyse, dont 74 utilisés dans l'évaluation 2016.

4.2.3.c Expression écrite

En considérant l'ensemble des items sur les trois années, il y avait au départ :

- 5 items communs aux trois années
- 16 items communs à 2004 et 2010
- 4 items communs à 2010 et 2016
- 12 items de 2004
- 5 items de 2010
- 66 items de 2016

Cela représente 108 items passés par les élèves en tout, dont 75 en 2016.

Après suppression des items présentant un mauvais Rbis, un fonctionnement différentiel ou un mauvais ajustement, il reste :

- 4 items communs aux trois années
- 13 items communs à 2004 et 2010
- 0 items communs à 2010 et 2016
- 12 items de 2004
- 4 items de 2010
- 66 items de 2016

104 items sont donc conservés dans l'analyse, dont 70 utilisés dans l'évaluation 2016.

4.3 Calcul des scores

Comme indiqué précédemment, une analyse conjointe des données (2004, 2010 et 2016) a permis d'estimer les paramètres des items, puis les niveaux de compétences θ des élèves. Afin de lever l'indétermination du modèle, la moyenne des θ a été fixé à 250 et leur écart-type à 50, pour l'échantillon de 2004. Le tableau 17 présente les résultats obtenus.

Tableau 17 – Niveaux de compétences CEDRE Anglais École (moyennes et écarts-type)

	Année	Nb élèves	Moyenne	Ecart-Type
Compréhension de l'écrit	2004	5989	250.0	50.0
	2010	6014	272.4	54.2
	2016	6014	277.3	51.0
Compréhension de l'oral	2004	5942	250	50
	2010	5988	260.7	53.5
	2016	5988	257.5	47.7
Expression écrite	2004	5944	250	50
	2010	6014	265.1	51.4
	2016	6014	259.5	47.3

5 Construction de l'échelle

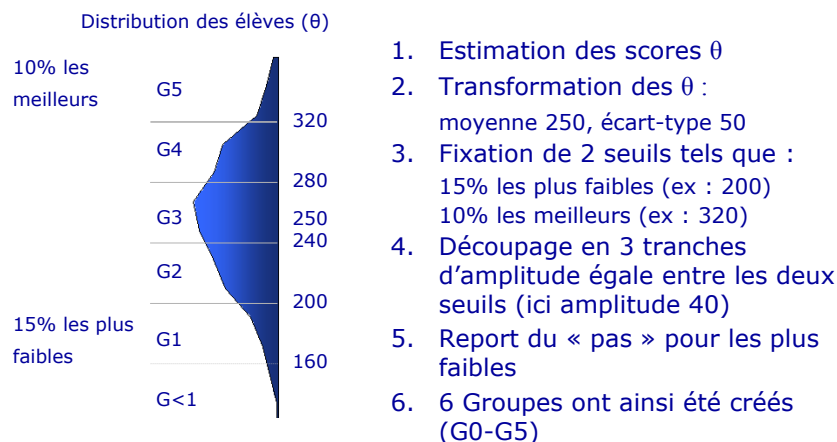
5.1 Méthode

Les modèles de réponse à l'item permettent de positionner sur une même échelle les paramètres de difficulté des items et les niveaux de compétences des élèves. Cette correspondance permet de caractériser les compétences maîtrisées pour différents groupes d'élèves.

Les scores en Anglais estimés selon le modèle de réponse à l'item présenté dans la partie précédente ont été standardisés de manière à obtenir une moyenne de 250 et un écart-type de 50 pour l'année 2004. Puis, comme le montre la figure 6, la distribution des scores est « découpée » en six groupes de la manière suivante : nous déterminons le score-seuil en-deça duquel se situent 15 % des élèves (groupes < 1 et 1), nous déterminons le score-seuil au-delà duquel se situent 10 % des élèves (groupe 5). Entre ces deux niveaux, l'échelle a été scindée en trois parties d'amplitudes de scores égales correspondant à trois groupes intermédiaires. Ces choix sont arbitraires et ont pour objectif de décrire plus précisément le continuum de compétence.

En effet, les modèles de réponse à l'item ont l'avantage de positionner sur la même échelle les scores des élèves et les difficultés des items. Ainsi, chaque item est associé à un des six groupes, en fonction des probabilités estimées de réussite selon les groupes. Un item est dit « maîtrisé » par un groupe dès lors que l'élève ayant le score le plus faible du groupe a au moins 50 % de chance de réussir l'item. Les élèves du groupe ont alors plus de 50 % de chance de réussir cet item.

Figure 6 – Principes de construction de l'échelle



5.2 Caractérisation des groupes de niveaux

A partir de cette correspondance entre les items et les groupes, une description qualitative et synthétique des compétences maîtrisées par les élèves des différents groupes est proposée.

5.2.1 Compréhension de l'oral

Groupe < 1 (2 % des élèves)

Les élèves de ce groupe parviennent à effectuer quelques opérations de repérage et ils commencent à pouvoir identifier des mots isolés, avec toutefois une aide visuelle. Ils ont quelques acquis lexicaux, se limitant à des mots très simples.

Groupe 1 (11 % des élèves)

Ces élèves sont capables de reconnaître et d'identifier des lettres de l'alphabet ou des blocs lexicalisés. Ils comprennent des expressions familières courantes et ils commencent à pouvoir comprendre une courte description ou des dialogues simples.

Groupe 2 (26,1 % des élèves)

Ces élèves sont capables de repérages plus fins, en contexte. Ils commencent à pouvoir discriminer des sons proches. Ils comprennent l'heure et les nombres à l'oral et ils peuvent repérer des informations explicites dès lors que celles-ci s'appuient sur du lexique simple (relatif à la famille, les animaux, les sports par exemple).

Groupe 3 (28,3 % des élèves)

Pour les élèves de ce groupe, la compréhension s'affine grâce à l'acquisition d'un lexique plus étendu. Ils peuvent associer questions et réponses usuelles dans le contexte du rituel de classe. Ils sont maintenant capables de mobiliser des indices grammaticaux, lexicaux et phonologiques plus variés et plus nombreux, même sans aide visuelle. Ils commencent à pouvoir les mettre en réseau pour construire le sens.

Groupe 4 (17,9 % des élèves)

Les compétences évaluées sont bien installées. Les élèves progressent dans la discrimination au sein d'un continuum sonore plus dense et plus long. Leurs acquis linguistiques sont plus solides, ce qui facilite leur accès au sens. Ils peuvent comprendre des descriptions et de courts dialogues. Ils commencent à accéder à l'implicite dans un document sonore.

Groupe 5 (14,7 % des élèves)

Les élèves de ce groupe ont des compétences affirmées en compréhension de l'oral. Ils peuvent s'appuyer sur des connaissances solides et mobiliser des capacités telles qu'identifier les mots porteurs de sens, dégager les principales informations d'un message sonore et construire du sens en mettant en relation plusieurs éléments, à partir de supports variés. Les élèves sont capables de synthétiser l'information, compétence inaccessible aux groupes inférieurs.

5.2.2 Compréhension de l'écrit**Groupe < 1 (0,3 % des élèves)**

Bien que capables de répondre ponctuellement à quelques questions, les élèves ne maîtrisent quasiment aucune des compétences attendues en fin d'école.

Groupe 1 (3,5 % des élèves)

Les élèves ont une connaissance limitée du lexique de base relatif à leur environnement proche. Ils sont capables ponctuellement d'identifier certains éléments d'information pour peu que ceux-ci soient très facilement identifiables et le lexique transparent.

Groupe 2 (17,9 % des élèves)

Les élèves de ce groupe parviennent à prélever ou à discriminer une information à partir d'indices lexicaux explicites ou en s'appuyant sur une illustration. Ils connaissent des éléments de lexique courants, touchant à leur sphère immédiate (famille, vêtements, animaux). Ils reconnaissent aisément les mots transparents dans un support écrit.

Groupe 3 (32,2 % des élèves)

Ces élèves ont une bonne connaissance du lexique de base et des expressions courantes. Ils peuvent dégager la principale information d'un texte. Dans des textes plus longs, même sans aide visuelle, ils sont capables de prélever des indices moins facilement repérables et de commencer à opérer une mise en relation.

Groupe 4 (24,3 % des élèves)

Ces élèves ont une maîtrise satisfaisante du lexique de base et des expressions courantes, attendus en fin d'école. Ils sont capables de comprendre un énoncé à partir d'éléments implicites, sans l'aide d'un support iconographique. Ils peuvent s'appuyer sur des outils linguistiques plus complexes pour construire du sens et synthétiser.

Groupe 5 (21,8 % des élèves)

Ce groupe d'élèves a un niveau de performance très satisfaisant dans chacune des compétences évaluées. Ces élèves peuvent comprendre des textes courts et simples de natures diverses (textes informatifs, narratifs, descriptifs) et ils ont la capacité d'en dégager les principales informations, mais aussi d'accéder à l'implicite. Ils sont capables de comprendre des messages et des descriptions et ils peuvent synthétiser l'information.

5.3 Exemples d'items

5.3.1 Compréhension de l'oral

5.3.1.a Item caractéristique du groupe < 1

Script

Jenny part en classe de mer quelques jours avec sa classe. La maîtresse rappelle aux élèves ce qu'ils doivent mettre dans la valise.

La liste va être entendue deux fois.

Coche la liste qui correspond aux vêtements demandés par la maîtresse.

-Shorts

-A pair of jeans

-Three tee shirts

-shoes

Figure 7 – Exemple groupe < à 1



Cette situation, qui évalue la maîtrise de la compétence " Connaître et recon-

naître des termes lexicaux de base », propose aux élèves un contexte familier : le départ en classe de mer. Ils entendent une liste de vêtements, qu'ils doivent identifier en l'associant au visuel correspondant. Le niveau lexical est très simple et très accessible pour des élèves de fin d'école primaire : shorts, jeans, tee shirts, shoes, ce qui explique sans doute le fort taux global de réussite : 88,5 %. Le lexique usuel des vêtements et accessoires courants est maîtrisé dans cet item. Les réponses 1 et 2 (erronées) sont choisies respectivement par 2,7 % et 8,3 % des élèves. La première " valise " contenait bien un short et un jean (à noter le pluriel anglais shorts et jeans, correspondant à " un short " et " un jean " en français) mais seulement deux t-shirts et des bottes. La proximité phonologique entre " boots " et " shoes " a pu induire certains élèves en erreur mais c'est surtout la non-identification du chiffre trois (" three "), très proche lui aussi phoniquement du " tee " de t-shirts, qui peut expliquer le choix de cette réponse. Dans la réponse 2, qui propose bien un short, 3 t-shirts et des chaussures, c'est l'item " a pair of jeans " qui fait la différence et révèle la difficulté pour 8 % de l'échantillon de segmenter le bloc sonore " a pair of jeans " et d'isoler " jeans " après " of ".

5.3.1.b Item caractéristique du groupe 1

Script

-Mum, I need 2 pairs of jeans, white tee-shirts and a black jacket.

-How many tee shirts ?

-3 white tee shirts. Oh, and a pair of shoes.

-What colour for your shoes ?

-I would like brown shoes.

Figure 8 – Exemple groupe 1

Bob et sa mère vont faire des courses. Lis les questions avant d'écouter le dialogue.
Tu vas entendre ce dialogue deux fois.

Dans chaque partie, coche les propositions qui conviennent.

Bob a besoin ...

- 1 de jeux.
- 2 de fournitures d'école.
- 3 de nourriture.
- 4 de vêtements.

ES4

Cette situation représentative des performances du groupe 1 évalue la maîtrise de

Figure 9 – Exemple groupe 2

Bethany va se présenter. Lis chaque affirmation et coche "VRAI" ou "FAUX".

	Vrai	Faux
Bethany a 10 ans et demi.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
Bethany fête son anniversaire en avril.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
Bethany a 2 chiens.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
Bethany a 2 sœurs.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
Bethany adore jouer au football.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂

la compétence "dégager les principales informations dans un document sonore" et plus particulièrement "comprendre un court dialogue". Il s'agit pour les élèves d'écouter une brève conversation entre un garçon et sa mère, qui s'apprêtent à effectuer des achats. Les élèves doivent tout d'abord repérer à l'oral et identifier les mots porteurs de sens et d'accent (jeans, tee-shirts, jacket, shoes) puis les mettre en relation afin de trouver le champ lexical : ici, les vêtements. Cette situation ne pose pas de difficulté aux élèves, ainsi qu'en témoigne le taux global de réussite de 86,1 %. Ils savent repérer des items lexicaux familiers au sein de la chaîne sonore (jeans et tee-shirts notamment), items qui sont par ailleurs utilisés à l'identique en français dans le langage courant. Il est à noter que la seconde partie de cette situation propose des questions plus précises sur le document sonore. La question : "Bob voudrait quelque chose de marron : de quoi s'agit-il ?" réussie à 67,8 % pose davantage problème que celle présentée plus haut : elle est représentative du groupe 3.

5.3.1.c Item caractéristique du groupe 2

Script

Hello! My name is Bethany. I'm ten and a half years old and my birthday is on the 24th of April. I have green eyes and long blond hair. I have four brothers and no sisters. I have three dogs. Their names are Sage, Pepper and Cowboy. My favourite hobby is to draw and my favourite sport is basketball. I do not like rugby.

My favourite food is pizza and I do not like salad.

Dans cette situation, caractéristique du groupe 2, les élèves entendent une personne se présenter et doivent se prononcer sur l'exactitude de plusieurs affirmations, en cochant vrai ou faux. L'objectif d'évaluation visé ici est " Comprendre une brève description de quelqu'un ", qui s'inscrit lui-même dans le cadre plus général de " Dégager les principales informations d'un document sonore ". Les items 1, 3, 4 et 5, dont les taux globaux de réussite se situent entre 75,8 % et 80 %, indiquent une maîtrise satisfaisante de la capacité à comprendre les informations-clés dans un message sonore. Huit élèves sur dix savent identifier l'âge de Bethany au sein de la chaîne sonore, ce qui reflète bien leurs compétences en construction. En effet, la forme complexe " ten and a half years old " ne présente pas pour eux de difficulté majeure, pour être sans doute pratiquée régulièrement dans les classes. La notion de possession est maîtrisée elle aussi, avec un taux global de réussite de 80 % : la forme verbale " I have ", associée aux animaux domestiques (" I have three dogs ") et aux membres de la famille (" I have four brothers and no sisters ") est elle aussi bien ancrée dans les pratiques de classe à l'école primaire. L'item 2, réussi à 60 %, pose davantage problème : il s'agit d'identifier la date d'anniversaire de Bethany : the 24th of April. Les échanges autour de la thématique des anniversaires font partie des rituels de classe, auxquels les élèves de primaire s'adonnent joyeusement. La date est souvent dite à chaque début de séance et inscrite au tableau. Dans ces conditions, le taux de réussite plus faible que celui des réponses 1, 3, 4 et 5 pose question et soulève le problème de la reconnaissance d'un lexique dit " connu " dans le contexte spécifique de la compréhension de l'oral. En effet, les élèves, qui ont peine à reconnaître le mois d'avril dans le flux sonore, n'ont pu associer ce mois qu'ils connaissent pourtant bien, à sa réalisation phonique, peut-être éloignée de leur propre prononciation de ce mot. Sont en jeu ici des compétences de discrimination auditive plus fine, accessibles aux élèves d'un groupe de performance supérieur, le groupe 3.

5.3.1.d Item caractéristique du groupe 3

Script

Hello Jenny.

Hi Bob.

Do you like swimming Jenny ?

Yes, I do but I prefer football. Do you like football Bob ?

No, I don't.

What's your favourite sport ?

My favourite sport is running.

Figure 10 – Exemple groupe 3

**Deux enfants Bob et Jenny attendent pour s'inscrire au centre sportif.
Tu vas entendre leur conversation deux fois.
Coche la réponse qui convient pour chaque question.**

Bob aime-t-il le football?

- 1 Oui
- 2 Non
- 3 On ne sait pas.

Quel est le sport préféré de Bob?

- 1 La natation
- 2 La course à pied
- 3 Le football

Cette situation représentative du groupe 3 évalue la maîtrise de la compétence " Comprendre un court dialogue " (Dégager les principales informations dans un document sonore). Les élèves écoutent un échange entre deux personnages et doivent ensuite répondre à des questions par QCM portant sur le personnage Bob. La question 1 ne pose pas de difficulté majeure aux élèves qui identifient correctement à travers la réponse brève de Bob " No, I don't " le fait qu'il n'aime pas le football (" Do you like football Bob ? "). Cet item est réussi à 73,3%. La question 2, portant sur le sport préféré du personnage, est moins bien réussie (62,8%). La fin du dialogue présente en effet quelques difficultés : croisement des réponses des personnages masculin et féminin, différenciation de " like " et de " favourite ", identification du mot " running ", proximité phonique de " swimming " et de " running ". Il met en lumière les limites des compétences maîtrisées par les élèves de ce groupe de performances.

5.3.1.e Item caractéristique du groupe 4

Script

Hello Pierre. I'm very happy to see you!

Hi, Mary, welcome to France.

I'm excited to meet your family.

Here are my parents, Paul and Nathalie. And this is my brother Enzo.
 Here's a present for you.
 Oh, lovely! Chocolate! We like eating chocolate?

Figure 11 – Exemple groupe 4

Ecoute ce dialogue entre deux enfants, puis pour chaque question, coche la proposition qui convient.

❶ Pierre parle :

- 1 avec sa maîtresse
 2 avec ses parents
 3 avec sa correspondante
 4 avec sa voisine

❷ Mary apporte :

- 1 
- 2 
- 3 
- 4 

Cette situation représentative du groupe 4 se propose d'évaluer la compétence "comprendre un court dialogue" (Dégager les principales informations d'un document sonore) dont la thématique est l'accueil d'un correspondant étranger. Le jeune Pierre souhaite la bienvenue à sa correspondante britannique Mary et lui présente sa famille : ses parents et son frère Enzo. La première question ("Pierre parle avec??") permet de mesurer l'aptitude des élèves à synthétiser. Il leur faut en effet recueillir les différents indices présents dans le document : indices linguistiques (welcome, meet your family, excited), pragmatiques (codes socio-culturels évoqués par "welcome to France") mais aussi paralinguistiques (intonations exprimant l'enthousiasme d'une première rencontre, accents français et britanniques). Puis il convenait de les mettre en relation afin d'aboutir à

une hypothèse conclusive pertinente : Pierre parle bien avec sa correspondante. Le taux global de réussite de l'item 1 est révélateur de la complexité des stratégies à mettre en œuvre ici pour parvenir au sens : 49 % soit à peine un élève sur deux. L'accès à l'implicite constitue encore une difficulté importante pour les élèves de l'école primaire, qui ne parviennent pas à ce stade à déduire le sens à partir d'éléments explicites.

5.3.1.f Item caractéristique du groupe 5

John a une conversation avec sa mère.

Ecoute le dialogue.

Coche " vrai ", " faux " ou " ce n'est pas précisé ".

Script

- Come to your bed, John.
- Why ?
- It's time to sleep.
- I don't want to sleep mum.
- It's nine o'clock.
- I'm not tired.
- There's school tomorrow. You have to sleep now. Good night John.
- Good night mum.

Figure 12 – Exemple groupe 5

	vrai	faux	Ce n'est pas précisé
1. John est très fatigué.			
2. Il est neuf heures du soir.			
3. Le lendemain, il doit aller à l'école.			
4. Sa mère lui raconte toujours une histoire avant d'aller se coucher.			

Cette situation représentative des performances du groupe 5 évalue la composante " Comprendre un court dialogue " au sein de la compétence " Dégager les principales informations dans un message sonore ". Les élèves écoutent un dialogue et doivent se prononcer sur l'exactitude des affirmations proposés : sont-elles vraies, fausses ou l'information n'est-elle tout simplement pas présente dans le document ? Les items 2 et 3 relèvent des performances du groupe 3 et sont réussis respectivement à 63,7 % et 68,2 %. L'item 1, quant à lui, n'est accessible qu'à partir du groupe 4 : les élèves peinent à identifier la négation

dans " I'm not tired " et à opérer une segmentation entre les différents mots, notamment " not " et " tired " et ils hésitent nettement entre " vrai " (réponse choisie à 30,1 %) et " faux " (43,8 %). 12,3 % des élèves évalués optent pour " ce n'est pas précisé ", ce qui révèle leur incapacité à repérer et identifier le mot " tired ". L'item 5 " Sa mère lui raconte toujours une histoire avant d'aller se coucher " pose des problèmes encore plus marqués, puisqu'il ne s'agit pas ici de repérer un item lexical, fût-il inclus dans une phrase négative, mais de comprendre l'ensemble du message pour être en capacité d'affirmer : " cette information n'est pas précisée dans l'enregistrement ". Seuls 35,5 % des élèves se sont trouvés en réussite, ce qui place cet item à la portée du seul groupe 5 de l'échantillon.

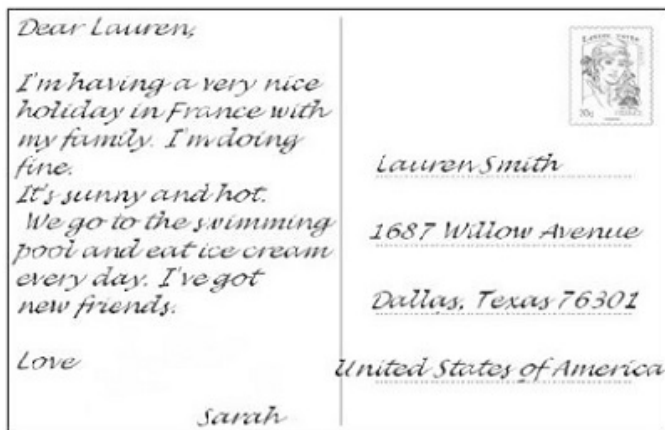
5.3.2 Compréhension de l'écrit

5.3.2.a Item caractéristique du groupe 1

La situation présentée ici pour le groupe 1 évalue la compétence " dégager les principales informations dans une carte postale " mais seule la question 3 (" Elle va bien ") relève des performances de ce groupe. Les élèves parviennent aisément à comprendre à travers ses propos écrits que tout se passe bien pour la jeune Sarah durant son séjour en France. Le taux global de réussite est en effet de 87,8 % et indique que le bloc lexicalisé " I'm (doing) fine " est correctement identifié. Pratiqué en classe de façon rituelle, dans le cadre des relations de civilité (" How are you today ? " " I'm (doing) fine "), il ne constitue pas de difficulté, même si les élèves doivent parvenir tout d'abord à l'isoler pour pouvoir l'identifier, lors de leur lecture du document.

Figure 13 – Exemple groupe 1

Sarah envoie une carte postale à son amie Lauren.



Lis la carte postale. Puis pour chacune des affirmations suivantes, coche Vrai ou Faux.

	Vrai	Faux
Sarah habite en France.	<input type="checkbox"/> 1	<input type="checkbox"/> 2
Il pleut chaque jour.	<input type="checkbox"/> 1	<input type="checkbox"/> 2
Elle va bien.	<input type="checkbox"/> 1	<input type="checkbox"/> 2

5.3.2.b Item caractéristique du groupe 2

La situation présentée ici pour le groupe 1 évalue la compétence "dégager les principales informations dans une carte postale" mais seule la question 3 ("Elle va bien") relève des performances de ce groupe. Les élèves parviennent aisément à comprendre à travers ses propos écrits que tout se passe bien pour la jeune Sarah durant son séjour en France. Le taux global de réussite est en effet de 87,8 % et indique que le bloc lexicalisé "I'm (doing) fine" est correctement identifié. Pratiqué en classe de façon rituelle, dans le cadre des relations de civilité ("How are you today?" "I'm (doing) fine"), il ne constitue pas de difficulté, même si les élèves doivent parvenir tout d'abord à l'isoler pour pouvoir l'identifier, lors de leur lecture du document.

Figure 14 – Exemple groupe 2

Vrai ou faux ? Observe le dessin. Pour chaque proposition, coche "Vrai" ou "Faux".



	Vrai	Faux
Mary is in the garden.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
She has got a cat.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
She is reading a book.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
She is wearing a dress.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
She has got long hair.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂
We can see the British flag.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂

5.3.2.c Item caractéristique du groupe 3

La situation présentée ici pour le groupe 3 est conçue pour mesurer la capacité des élèves à se faire une idée du contenu d'un texte en identifiant le type de document. On propose quatre types de document différents : une recette de cuisine, des horaires de train, une notice de montage et une lettre. Les taux globaux de réussite se situent entre 74 % et 89 %, ce qui souligne une maîtrise satisfaisante des éléments à la fois linguistiques et paralinguistiques présents dans les supports. Les élèves de ce groupe peuvent construire du sens à partir d'éléments épars, qui relèvent à la fois de la langue (" fruit salad " ou " instructions "), du paralinguistique (la mise en page ou la disposition d'éléments visuels ou textuels) ainsi que de leur expérience du monde. En effet, ils peuvent également faire appel à leurs connaissances pratiques, à leur expérience dans la vie quotidienne et s'en servir pour les transférer dans un contexte de langue étrangère. Ces stratégies, plus complexes et fines, sont révélatrices des performances du groupe 3.

Figure 15 – Exemple groupe 3

<p style="text-align: center;">Fruit Salad Recipe</p> <p style="text-align: center;">2 Cups Sliced Peaches 2 Cups Sliced Strawberries 2 Cups Blueberries 2 Cups Sliced Grapes</p> <p style="text-align: center;">Topping 8 oz soft Cream Cheese 8 oz Sour Cream 1/2 cup Sugar 1/2 tsp Vanilla</p> <p style="text-align: center;">Mix together and beat until smooth</p> <p style="text-align: center;">Layer Fruit and ice with topping Top with Crushed Pecans if desired</p> <p style="text-align: center;">A</p>	<table border="1"> <thead> <tr> <th colspan="2">Saturday</th> <th colspan="8">London Waterloo - Portsmouth - Isle of Wight</th> </tr> <tr> <th></th> <th>Times</th> <th>==</th> <th>==</th> <th>==</th> <th>==</th> <th>==</th> <th>==</th> <th>==</th> <th>==</th> </tr> </thead> <tbody> <tr> <td>London Waterloo</td> <td>08:00</td> <td>10:30</td> <td>13:00</td> <td>15:30</td> <td>18:00</td> <td>19:30</td> <td>17:30</td> <td>17:30</td> <td>18:30</td> </tr> <tr> <td>Woking</td> <td>08:40</td> <td>10:55</td> <td>13:25</td> <td>15:55</td> <td>18:25</td> <td>19:55</td> <td>17:25</td> <td>17:55</td> <td>18:55</td> </tr> <tr> <td>Guildford</td> <td>09:20</td> <td>11:35</td> <td>14:05</td> <td>16:35</td> <td>19:05</td> <td>20:35</td> <td>17:35</td> <td>18:05</td> <td>19:05</td> </tr> <tr> <td>Haslemere</td> <td>09:55</td> <td>12:10</td> <td>14:40</td> <td>17:10</td> <td>19:40</td> <td>21:10</td> <td>17:50</td> <td>18:20</td> <td>19:20</td> </tr> <tr> <td>Portsmouth</td> <td>10:30</td> <td>12:45</td> <td>15:15</td> <td>17:45</td> <td>20:15</td> <td>21:45</td> <td>18:30</td> <td>19:00</td> <td>20:00</td> </tr> <tr> <td>Portsmouth Harbour</td> <td>10:55</td> <td>13:10</td> <td>15:40</td> <td>18:10</td> <td>20:40</td> <td>22:10</td> <td>18:55</td> <td>19:25</td> <td>20:25</td> </tr> <tr> <td>Portsmouth Harbour</td> <td>15:40</td> <td>17:15</td> <td>17:10</td> <td>18:15</td> <td>18:45</td> <td>19:15</td> <td>19:15</td> <td>20:15</td> <td>21:15</td> </tr> <tr> <td>Ryde Pier Head</td> <td>16:20</td> <td>17:20</td> <td>18:20</td> <td>19:20</td> <td>19:50</td> <td>20:20</td> <td>20:20</td> <td>21:20</td> <td>22:20</td> </tr> <tr> <td>Ryde Pier Head</td> <td>17:30</td> <td>17:45</td> <td>18:15</td> <td>18:45</td> <td>19:15</td> <td>19:45</td> <td>19:45</td> <td>20:15</td> <td>20:45</td> </tr> <tr> <td>Ryde Esplanade</td> <td>17:55</td> <td>18:10</td> <td>18:40</td> <td>19:10</td> <td>19:40</td> <td>20:10</td> <td>20:10</td> <td>20:40</td> <td>21:10</td> </tr> <tr> <td>Ryde St Johns Road</td> <td>18:10</td> <td>18:25</td> <td>18:55</td> <td>19:25</td> <td>19:55</td> <td>20:25</td> <td>20:25</td> <td>20:55</td> <td>21:25</td> </tr> <tr> <td>Smallbrook Junction</td> <td>18:25</td> <td>18:40</td> <td>19:10</td> <td>19:40</td> <td>20:10</td> <td>20:40</td> <td>20:40</td> <td>21:10</td> <td>21:40</td> </tr> <tr> <td>Basingstoke</td> <td>18:40</td> <td>18:55</td> <td>19:25</td> <td>19:55</td> <td>20:25</td> <td>20:55</td> <td>20:55</td> <td>21:25</td> <td>21:55</td> </tr> <tr> <td>Sandown</td> <td>18:55</td> <td>19:10</td> <td>19:40</td> <td>20:10</td> <td>20:40</td> <td>21:10</td> <td>21:10</td> <td>21:40</td> <td>22:10</td> </tr> <tr> <td>Lake</td> <td>19:10</td> <td>19:25</td> <td>19:55</td> <td>20:25</td> <td>20:55</td> <td>21:25</td> <td>21:25</td> <td>21:55</td> <td>22:25</td> </tr> <tr> <td>Shanklin</td> <td>19:25</td> <td>19:40</td> <td>20:10</td> <td>20:40</td> <td>21:10</td> <td>21:40</td> <td>21:40</td> <td>22:10</td> <td>22:40</td> </tr> <tr> <td>Ventnor (Boats store)</td> <td>19:40</td> <td>19:55</td> <td>20:25</td> <td>20:55</td> <td>21:25</td> <td>21:55</td> <td>21:55</td> <td>22:25</td> <td>22:55</td> </tr> </tbody> </table> <p style="text-align: center;">B</p>	Saturday		London Waterloo - Portsmouth - Isle of Wight									Times	==	==	==	==	==	==	==	==	London Waterloo	08:00	10:30	13:00	15:30	18:00	19:30	17:30	17:30	18:30	Woking	08:40	10:55	13:25	15:55	18:25	19:55	17:25	17:55	18:55	Guildford	09:20	11:35	14:05	16:35	19:05	20:35	17:35	18:05	19:05	Haslemere	09:55	12:10	14:40	17:10	19:40	21:10	17:50	18:20	19:20	Portsmouth	10:30	12:45	15:15	17:45	20:15	21:45	18:30	19:00	20:00	Portsmouth Harbour	10:55	13:10	15:40	18:10	20:40	22:10	18:55	19:25	20:25	Portsmouth Harbour	15:40	17:15	17:10	18:15	18:45	19:15	19:15	20:15	21:15	Ryde Pier Head	16:20	17:20	18:20	19:20	19:50	20:20	20:20	21:20	22:20	Ryde Pier Head	17:30	17:45	18:15	18:45	19:15	19:45	19:45	20:15	20:45	Ryde Esplanade	17:55	18:10	18:40	19:10	19:40	20:10	20:10	20:40	21:10	Ryde St Johns Road	18:10	18:25	18:55	19:25	19:55	20:25	20:25	20:55	21:25	Smallbrook Junction	18:25	18:40	19:10	19:40	20:10	20:40	20:40	21:10	21:40	Basingstoke	18:40	18:55	19:25	19:55	20:25	20:55	20:55	21:25	21:55	Sandown	18:55	19:10	19:40	20:10	20:40	21:10	21:10	21:40	22:10	Lake	19:10	19:25	19:55	20:25	20:55	21:25	21:25	21:55	22:25	Shanklin	19:25	19:40	20:10	20:40	21:10	21:40	21:40	22:10	22:40	Ventnor (Boats store)	19:40	19:55	20:25	20:55	21:25	21:55	21:55	22:25	22:55
Saturday		London Waterloo - Portsmouth - Isle of Wight																																																																																																																																																																																													
	Times	==	==	==	==	==	==	==	==																																																																																																																																																																																						
London Waterloo	08:00	10:30	13:00	15:30	18:00	19:30	17:30	17:30	18:30																																																																																																																																																																																						
Woking	08:40	10:55	13:25	15:55	18:25	19:55	17:25	17:55	18:55																																																																																																																																																																																						
Guildford	09:20	11:35	14:05	16:35	19:05	20:35	17:35	18:05	19:05																																																																																																																																																																																						
Haslemere	09:55	12:10	14:40	17:10	19:40	21:10	17:50	18:20	19:20																																																																																																																																																																																						
Portsmouth	10:30	12:45	15:15	17:45	20:15	21:45	18:30	19:00	20:00																																																																																																																																																																																						
Portsmouth Harbour	10:55	13:10	15:40	18:10	20:40	22:10	18:55	19:25	20:25																																																																																																																																																																																						
Portsmouth Harbour	15:40	17:15	17:10	18:15	18:45	19:15	19:15	20:15	21:15																																																																																																																																																																																						
Ryde Pier Head	16:20	17:20	18:20	19:20	19:50	20:20	20:20	21:20	22:20																																																																																																																																																																																						
Ryde Pier Head	17:30	17:45	18:15	18:45	19:15	19:45	19:45	20:15	20:45																																																																																																																																																																																						
Ryde Esplanade	17:55	18:10	18:40	19:10	19:40	20:10	20:10	20:40	21:10																																																																																																																																																																																						
Ryde St Johns Road	18:10	18:25	18:55	19:25	19:55	20:25	20:25	20:55	21:25																																																																																																																																																																																						
Smallbrook Junction	18:25	18:40	19:10	19:40	20:10	20:40	20:40	21:10	21:40																																																																																																																																																																																						
Basingstoke	18:40	18:55	19:25	19:55	20:25	20:55	20:55	21:25	21:55																																																																																																																																																																																						
Sandown	18:55	19:10	19:40	20:10	20:40	21:10	21:10	21:40	22:10																																																																																																																																																																																						
Lake	19:10	19:25	19:55	20:25	20:55	21:25	21:25	21:55	22:25																																																																																																																																																																																						
Shanklin	19:25	19:40	20:10	20:40	21:10	21:40	21:40	22:10	22:40																																																																																																																																																																																						
Ventnor (Boats store)	19:40	19:55	20:25	20:55	21:25	21:55	21:55	22:25	22:55																																																																																																																																																																																						
<p style="text-align: center;">BASE BLIND CABINET ASSEMBLY INSTRUCTIONS</p> <p>Step 1- 1. Lay cabinet Back Panel (H) on a soft working surface with grooves facing upward. The grooves will be for the Bottom Panel (D).</p> <p>2. Position the Left and Right Side Panels (A) to be fitted with the top of the Back Panel (H) making sure clips are lined up and fully engaged. Insert locking screen into each clip pair and tighten all screws.</p> <p>Step 2- 1. Insert Bottom Panel (D) into the grooves of Side Panels (A) and Back Panel (H).</p> <p>Step 3- 1. Attach Front Face (B) onto Bottom Panel (D) and Side Panels (A) making sure panels fit into the grooves of the Front Face (B) and clips are lined up and fully engaged. Insert locking screen into each clip pair and tighten all screws.</p> <p>2. Attach Top Kick Plate (C) to Side Panels (A) making sure clips are lined up and fully engaged. Insert locking screen into each clip pair and tighten all screws.</p> <p style="text-align: center;">C</p>	<p style="text-align: center;">East Witness 10/10/2012 www.1010.org</p> <p>His Excellency László Kőrösi Speaker of the Hungarian National Assembly Hungarian National Assembly Kossuth tér 1-3 H-1052 Budapest Hungary</p> <p>June 7, 2012</p> <p>Mr. Speaker:</p> <p>It is with profound honor and indignation that I invited of your participation, together with Hungarian Secretary of State for Culture, Education and Religion, János László, to attend the ceremony in Romania honoring Grand Marshal, a member of the National Socialist Arrow Cross Parliament. I found it outrageous that the Speaker of the Hungarian National Assembly could participate in a ceremony honoring a Hungarian fascist ideologue of the Heiligh and Steiner regimes. This disturbing news came following the outrageous practice of naming public spaces after wartime leader Miklós Horthy and/or celebrating other fascist and other figures who collaborated heavily with the Hungarian fascist regime. I was also informed that the writings of extreme right ideologues are systematically introduced in the Hungarian curriculum.</p> <p style="text-align: center;">D</p>																																																																																																																																																																																														

Figure 16 – Exemple groupe 3

	Lettre	Programme de télévision	Horaires de trains	Notice de montage	Recette de cuisine
Document A	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
Document B	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
Document C	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
Document D	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅

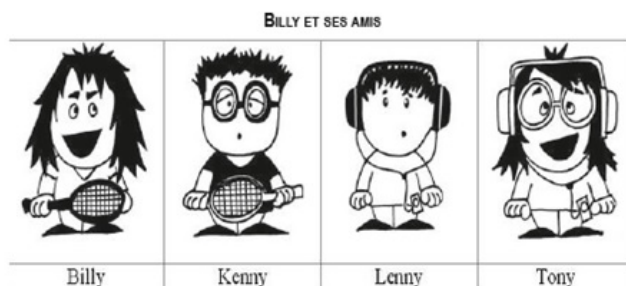
5.3.2.d Item caractéristique des groupes 4 et 5

Cette situation représentative du groupe 4 s'inscrit dans l'évaluation de la compétence : " Dégager les principales informations dans un texte descriptif (brève description d'un lieu, d'une personne, d'un objet) ". Les élèves doivent associer les dessins représentant différents personnages à chacun des textes descriptifs. Il est à noter que l'item 1 (description de Lenny) relève du groupe 5. Nous y reviendrons plus loin. Les items 2 et 3, réussis respectivement à 62,3 % et 54,1 %, sont accessibles au groupe 4, mais on note de nombreuses hésitations parmi les élèves. La description 2 est attribuée à Billy à 27,9 % et celle de Tony est attribuée par 34 % des élèves à Lenny. Pour réussir ces items, il faut à la fois mettre en œuvre des stratégies de repérage, d'identification mais aussi de mise en relation et de synthèse. Or, pour les descriptions 2 et 3, presque 3 élèves sur 10 ont limité leurs investigations au seul repérage des mots transparents " tennis " et " music ", sans opérer de véritable croisement d'informations. Ces résultats soulignent la difficulté persistante, à l'école primaire, à opérer des discriminations fines et mettre en œuvre une démarche cohérente et surtout exhaustive d'accès au sens. Revenons sur l'item 1 qui propose une description du personnage Lenny. C'est l'item le moins réussi de cette situation, avec un taux global de réussite de 50 % et il est, sans surprise, représentatif du groupe 5. Dans cet item, les élèves lisent une description de personnage très proche des deux autres, à la différence près qu'elle contient un élément d'information clé, apporté par la forme verbale négative de have, conjuguée de surcroît à la 3ème personne du singulier : " he hasn't got glasses ". Peu maîtrisée et pratiquée à l'école primaire, cette forme peut sembler dérouter et mettre en difficulté les élèves les plus habiles de l'échelle. Mais on peut également penser que les acquis lexicaux ne sont pas encore suffisamment riches et diversifiés, même chez les élèves de ce groupe, qui hésitent nettement entre les personnages de Lenny et Tony (plébis-

cit  par 37,8 % des  l ves). Ils peinent   identifier les noms communs "glasses" (des lunettes) et "mouth" (la bouche) ainsi que les adjectifs "short" et "round". Tous ces mots, pratiqu s   l'oral dans des activit s de classe telles que les "nursery rhymes" ou des jeux comme "Qui est-ce?", peuvent d router les  l ves, plus entra n s   reconnaître leur forme orale qu' crite. Ce qui souligne l'importance toute particuli re et sp cifique   l'anglais, du lien essentiel entre la graphie et la phonie.

Figure 17 – Exemple groupes 4 et 5

Observe les personnages et lis les descriptions. Coche le pr nom correspondant   chaque description. Attention ! Il y a un portrait en trop.



	Billy	Kenny	Lenny	Tony
He is small. He has got short dark hair and a small mouth. He hasn't got glasses. He loves music.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
He is small. He has got short dark hair and a small mouth. He has got round glasses. He loves tennis.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
He is small. He has got long black hair. He has got round glasses. He loves music.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4

6 Variables contextuelles et non cognitives

6.1 Variables sociodémographiques et indice de position sociale

Un certain nombre de variables sociodémographiques permettent d'enrichir l'analyse des résultats. Le score moyen des élèves est ainsi analysé en fonction du genre, du retard scolaire et quand les effectifs le permettent en fonction du secteur d'enseignement. Le lecteur est invité à consulter la Note d'Information pour plus de détails (Dalibard & Marchois, 2017).

L'indice de position sociale mesure la proximité au système scolaire du milieu familial de l'enfant. Cet indice peut se substituer à la profession des parents pour mieux expliquer les parcours et la réussite scolaire de leurs enfants. Il consiste en une transformation des PCS en valeur numérique (Rocher, 2016).

Il n'a été possible d'établir des comparaisons qu'en termes de niveau social des écoles, et non au niveau individuel. En effet, en 2016, la PCS des parents est disponible pour chaque élève, mais elle ne l'était pas dans les cycles antérieurs. Pour chaque établissement des échantillons de 2004, 2010 et 2016, la moyenne de l'indice de position socio-scolaire a été calculée et la population a ensuite été découpée en quatre groupes selon les quartiles (tableau 18).

Tableau 18 – Score moyen selon l'indice de position sociale moyen de l'établissement (CEDRE anglais - Compréhension de l'écrit)

Indice moyen école	Année	Répartition (%)	Score moyen	Écart type
1er quart	2004	24.8	237	45
1er quart	2010	24.2	256	50
1er quart	2016	23.1	259	43
2e quart	2004	25.3	247	47
2e quart	2010	25.2	268	53
2e quart	2016	25.5	276	49
3e quart	2004	24.9	252	48
3e quart	2010	25.5	279	55
3e quart	2016	22.6	277	51
4e quart	2004	25.0	264	55
4e quart	2010	25.2	286	55
4e quart	2016	28.8	294	53

Note de lecture : en 2010, le score moyen des élèves appartenant au quart des écoles les plus défavorisées (1er quart) augmente de 19 points par rapport à 2004. Les évolutions significatives sont indiquées en gras.

Tableau 19 – Score moyen selon l'indice de position sociale moyen de l'établissement (CEDRE anglais - Expression écrite)

Indice moyen école	Année	Répartition (%)	Score moyen	Écart type
1er quart	2004	24.7	236	46
1er quart	2010	24.2	254	50
1er quart	2016	23.1	244	50
2e quart	2004	25.4	246.3	47
2e quart	2010	25.2	263	52
2e quart	2016	25.5	259	43
3e quart	2004	24.9	253	47
3e quart	2010	25.5	270	51
3e quart	2016	22.6	259	47
4e quart	2004	25.1	264	45.4
4e quart	2010	25.2	274	50
4e quart	2016	28.8	273	45

Note de lecture : en 2016, l'écart type des élèves appartenant au quart des écoles les plus défavorisées (1er quart) baisse de 10 points par rapport à 2010. Les évolutions significatives sont indiquées en gras.

Tableau 20 – Score moyen selon l'indice de position sociale moyen de l'établissement (CEDRE anglais - Compréhension de l'oral)

Indice moyen école	Année	Répartition (%)	Score moyen	Écart type
1er quart	2004	24.8	235	43
1er quart	2010	24.8	244	49
1er quart	2016	22.6	241	42
2e quart	2004	25.4	252	57
2e quart	2010	25.0	258	55
2e quart	2016	25.7	254	41
3e quart	2004	24.7	252	45
3e quart	2010	25.1	266	52
3e quart	2016	22.8	259	45
4e quart	2004	25.1	261	50
4e quart	2010	25.1	274	53
4e quart	2016	28.9	273	48

Note de lecture : en 2010, l'écart type des élèves appartenant au quart des écoles plus favorisées (4ème quart) augmente de 13 points par rapport à 2004. Les évolutions significatives sont indiquées en gras.

6.2 Élaboration des questionnaires de contexte

Pour pouvoir davantage enrichir l'analyse des résultats, deux questionnaires de contexte ont été élaborés. Un questionnaire élève a été ajouté à la fin du cahier d'évaluation et un questionnaire enseignant était adressé aux enseignants des classes participantes à l'évaluation. Ces questionnaires ont été élaborés en collaboration avec des chercheurs et des spécialistes en sciences de l'éducation.

Le questionnaire enseignant interroge les enseignants sur leur niveau de formation et leur ancienneté. Ce questionnaire inclut aussi des questions sur les pratiques pédagogiques, les stratégies d'enseignement, le sentiment d'efficacité personnelle etc.

Le questionnaire élève interroge des dimensions dites conatives intéressantes à mettre en lien avec le score obtenu à l'épreuve - les stratégies de lecture, la motivation, la perception de soi et l'anxiété scolaire. De plus, les élèves sont demandés d'évaluer la difficulté de l'épreuve et leur degré d'implication à faire le test.

Le questionnaire élève contient aussi un certain nombre de questions à renseigner par l'enseignant(e), il s'agit des questions concernant la catégorie socioprofes-

sionnelle des parents mais aussi le parcours de l'élève (raccourcissement de cycle ou maintien dans un cycle, orientation retenue etc.).

6.3 Motivation des élèves face à la situation d'évaluation

Les évaluations standardisées des élèves, telles que CEDRE ou PISA, renvoient à des enjeux politiques croissants, alors qu'elles restent à faible enjeu pour les élèves participants. Dans le système éducatif français, où la notation tient une place prépondérante, la question de la motivation des élèves face à ces évaluations mérite d'être posée.

Un instrument pour mesurer la motivation a été adapté à partir du « thermomètre d'effort » proposé dans PISA (Keskpaik. & Rocher, 2015). Cet instrument (cf. figure 18) a été introduit dans plusieurs évaluations conduites au niveau national par la DEPP, y compris dans CEDRE maîtrise de la langue. Les données recueillies permettent de distinguer la motivation de l'élève de la difficulté perçue du test, et ainsi de mieux appréhender le lien entre la motivation des élèves français et leur performance. L'analyse de ces données renseigne en outre sur le rôle de certaines caractéristiques, des élèves ou des évaluations elles-mêmes, dans le degré de motivation à répondre aux questions de l'évaluation.

Le tableau 21 présente les grands résultats de cet instrument.

Tableau 21 – Résultats de l'instrument de mesure de la motivation au test (CEDRE anglais école 2016)

		%
Comment as-tu trouvé les exercices de cette évaluation ?	Très faciles	11.6
	Faciles	55.5
	Difficiles	28.0
	Très difficiles	4.9
Je me suis bien appliqué(e) pour faire cette évaluation	Pas du tout d'accord	2.5
	Pas d'accord	8.6
	D'accord	53.1
Je me suis autant appliqué(e) pour faire cette évaluation que le travail quotidien de classe	Tout à fait d'accord	35.8
	Pas du tout d'accord	7.3
	Pas d'accord	20.8
	D'accord	39.7
	Tout à fait d'accord	32.2

Figure 18 – Instrument de mesure de la motivation au test

[Q1]

Comment as-tu trouvé les exercices de cette évaluation ?

- 1 Très faciles
 2 Faciles
 3 Difficiles
 4 Très difficiles

[Q2]

Es-tu d'accord avec ces affirmations ?

(Coche une case par ligne)

	Pas du tout d'accord	Pas d'accord	D'accord	Tout à fait d'accord
Je me suis bien appliqué(e) pour faire cette évaluation	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
Je me suis autant appliqué(e) à faire cette évaluation que le travail quotidien de classe	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4

7 Annexe

Certification AFNOR pour les évaluations CEDRE

La DEPP est engagée dans un processus de certification. Elle a obtenu en mars 2015 la certification pour les évaluations CEDRE.

Les finalités de la certification

Les finalités sont les suivantes :

- inscrire les processus d'évaluation dans une dynamique pérenne d'amélioration continue ;
- renforcer la prise en compte des attentes des usagers dans la formalisation des objectifs des évaluations et la restitution de leurs résultats ;
- faire reconnaître par une certification de service la qualité du service rendu et la continuité du respect des engagements pris.

Les enjeux pour la DEPP

Il y a deux enjeux forts pour la DEPP, l'un interne, l'autre externe :

- améliorer les processus de construction des instruments d'évaluation des acquis des élèves, fiabiliser ces processus par une démarche de contrôle-qualité ;
- valoriser l'enquête CEDRE comme un standard de qualité procédurale dans le domaine de l'évaluation.

Plus spécifiquement, le projet de certification des évaluations CEDRE est porteur d'enjeux pour la DEPP en termes de communication sur la validité scientifique, la sincérité, l'objectivité et la fiabilité des évaluations, ainsi que sur l'éthique et le professionnalisme des équipes.

La démarche qualité

Elle est fondée sur un référentiel élaboré sur mesure, selon une démarche officielle reconnue par les services publics et en lien avec les représentants des utilisateurs du service et les professionnels. La transparence vis-à-vis des usagers est assurée par la communication des résultats des enquêtes de satisfaction annuelles.

Les engagements de service

Le référentiel d'engagements comporte 18 engagements (cf. encadré page suivante).

Les engagements de service de la DEPP

Des objectifs clairs et partagés

Nous associons les parties intéressées à la définition de notre programme d'évaluation.

Nous formalisons dans un " cadre d'évaluation " les résultats attendus et les paramètres techniques de l'évaluation, ses délais et les limites associées aux moyens mis en œuvre.

Des évaluations fondées sur l'expertise pédagogique

Nous définissons avec les parties intéressées les acquis à évaluer et les mesurons en intégralité.

Nous mobilisons, tout au long de l'évaluation, un groupe expérimenté composé d'enseignants de terrain, de formateurs, d'inspecteurs et de chercheurs.

Tous nos items sont testés, analysés et validés avec le groupe expert avant d'être utilisés dans le cadre d'une évaluation.

Les meilleures pratiques méthodologiques et statistiques au service de l'objectivité

Afin de garantir l'application des meilleures méthodes statistiques, nous prenons en compte avec exigence les principes du " Code de bonnes pratiques de la statistique européenne ".

Nous tirons un échantillon représentatif garantissant le maximum de précision de mesure, à partir du plan de sondage défini dans le respect du " cadre d'évaluation ".

Nous garantissons l'objectivité et la qualité des données recueillies par la standardisation des processus d'administration et de correction des tests.

Une mesure fiable et des comparaisons temporelles pertinentes

Afin de garantir l'application des meilleures méthodes psychométriques, nous prenons en compte avec exigence les recommandations internationales sur l'utilisation des tests.

Nous analysons les réponses apportées par les élèves aux items afin d'en garantir la validité psychométrique.

Nous modélisons une échelle de compétences servant de référence et offrons des comparaisons temporelles fiables et lisibles.

Nous caractérisons les niveaux de cette échelle et déterminons avec le groupe expert les seuils de maîtrise des compétences évaluées, permettant de vous décrire en détail les performances des élèves.

Des analyses enrichies par des données de contexte

Nous systématisons le recueil d'informations standardisées relatives aux élèves et à leur environnement scolaire et social, dans le respect le plus strict des règles de confidentialité.

Nous éclairons les résultats de nos évaluations par la mise en relation des scores avec ces données.

Transparence des méthodes et partage des résultats

Nous publions et présentons les résultats de chacune de nos évaluations.

Nous mettons à disposition un rapport technique précisant les méthodes utilisées dans le cadre de l'évaluation.

Nous participons, dans le cadre de conventions collaboratives, à des analyses complémentaires des données que nous produisons.

Références

- Ardilly, P. (2006). *Les techniques de sondage*. Technip.
- Christine, M., & Rocher, T. (2012, janvier). Construction d'échantillons astreints à des conditions de recouvrement par rapport à un échantillon antérieur et à des conditions d'équilibrage par rapport à des variables courantes : aspects théoriques et mise en œuvre dans le cadre du renouvellement des échantillons des enquêtes d'évaluation des élèves. In *Journées de méthodologie statistique*. Paris.
- Dalibard, E., & Marchois, C. (2017). CEDRE 2004 - 2010 - 2016 - anglais et allemand en fin d'école : après une forte progression, des résultats désormais stabilisés. *Note d'information*, 19.
- Keskaik., S., & Rocher, T. (2015). La motivation des élèves français face à des évaluations à faibles enjeux. comment la mesurer ? son impact sur les réponses. *Education et formations*, 85-86, 119-139.
- Rocher, T. (1999). *Psychométrie et théorie des sondages* (Mémoire de Master non publié). Université Paris VI.
- Rocher, T. (2013). *Mesure des compétences : les méthodes se valent-elles ? questions de psychométrie dans le cadre de l'évaluation de la compréhension de l'écrit* (Thèse de doctorat non publiée). Université Paris-Ouest.
- Rocher, T. (2015). Mesure des compétences : méthodes psychométriques utilisées dans le cadre des évaluations des élèves. *Éducation et Formations*, 86-87, 37-60.
- Rocher, T. (2016). Construction d'un indice de position sociale des élèves. *Éducation et Formations*, 90, 5-27.
- Sautory, O. (1993). La macro calmar. redressement d'un échantillon par calage sur marges. *Série des documents de travail de l'INSEE, Document F9310*.
- Smith, R., Schumaker, R., & Bush, J. (1998). Using item mean squares to evaluate fit to the rasch model. *Journal of Outcome Measurement*, 2 n° 1, 66-78.
- Tillé, Y. (2001). *Théorie des sondages. échantillonnage et estimation en populations finies. cours et exercices avec solution*. Paris : Dunod.
- Trosseille, B., & Rocher, T. (2015). Les évaluations standardisées des élèves. perspective historique. *Éducation et Formations*, 85-86, 15-35.
- Warm, T. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54 n° 3, 427-450.

Liste des tableaux

1	Cycle des évaluations disciplinaires réalisées sur échantillons (CEDRE) depuis 2003	5
2	Définition des compétences en compréhension de l'oral (évaluation 2016)	7
3	Définition des compétences en compréhension de l'écrit (évaluation 2016)	8
4	Définition des compétences en expression écrite (évaluation 2016)	8
5	Définition des compétences en expression orale en continu (évaluation 2016)	8
6	Exclusions pour la base de sondage - CEDRE Anglais CM2 2016	17
7	Répartition base de sondage - CEDRE Anglais CM2 2016	17
8	Répartition dans l'échantillon - CEDRE Anglais CM2 2016	18
9	Non-réponse des écoles - CEDRE Anglais CM2 2016	19
10	Non-réponse globale - CEDRE Anglais CM2 2016	19
11	Comparaison entre les marges de l'échantillon et les marges dans la population - Ecrit	21
12	Comparaison entre les marges de l'échantillon et les marges dans la population - Oral	21
13	Scores moyens et erreurs standard associées - CEDRE Anglais CM2	22
14	Répartition en % dans les groupes de niveaux - CEDRE Anglais CM2	23
15	Erreurs standards des répartitions en % dans les groupes de niveaux - CEDRE Anglais CM2	23
16	Effet du plan de sondage - CEDRE Anglais CM2 2016- Compréhension de l'écrit	24
17	Niveaux de compétences CEDRE Anglais École (moyennes et écarts-type)	44
18	Score moyen selon l'indice de position sociale moyen de l'établissement (CEDRE anglais - Compréhension de l'écrit)	63
19	Score moyen selon l'indice de position sociale moyen de l'établissement (CEDRE anglais - Expression écrite)	64
20	Score moyen selon l'indice de position sociale moyen de l'établissement (CEDRE anglais - Compréhension de l'oral)	65
21	Résultats de l'instrument de mesure de la motivation au test (CEDRE anglais école 2016)	66

Table des figures

1	Représentation graphique utilisée pour le regroupement d'items .	30
2	Modèle de réponse à l'item - 2 paramètres	33

3	Exemples d'ajustements (FIT)	37
4	Comparaison des paramètres de difficulté 2004-2010 (CEDRE Anglais CM2)	40
5	Comparaison des paramètres de difficulté 2010-2016 (CEDRE Anglais CM2)	41
6	Principes de construction de l'échelle	46
7	Exemple groupe < à 1	49
8	Exemple groupe 1	50
9	Exemple groupe 2	51
10	Exemple groupe 3	53
11	Exemple groupe 4	54
12	Exemple groupe 5	55
13	Exemple groupe 1	57
14	Exemple groupe 2	58
15	Exemple groupe 3	59
16	Exemple groupe 3	60
17	Exemple groupes 4 et 5	61
18	Instrument de mesure de la motivation au test	67