

# **CEDRE**

Cycle des Évaluations Disciplinaires Réalisées sur Échantillons

## **Rapport technique**

Compétences langagières et littératie 2021

Collège

Auteurs :

Sandra ANDREU  
Hélène DURAND DE MONESTROL  
Yann ETEVE  
Marguerite GARNERO  
Louis PHILBERT  
Hugo ROGIE  
Ronan VOURC'H

Bureau de la conception et du pilotage des évaluations des élèves  
Bureau des études statistiques et psychométriques sur les évaluations des élèves  
DEPP - Direction de l'évaluation, de la prospective et de la performance  
Ministère de l'éducation nationale

Mars 2025

## Table des matières

<b>Introduction</b>	<b>3</b>
<b>1 Cadre d'évaluation</b>	<b>4</b>
1.1 Objectifs . . . . .	4
1.2 Connaissances et compétences visées . . . . .	4
1.3 Construction du test . . . . .	9
1.4 Présentation générale de la passation . . . . .	19
<b>2 Sondage</b>	<b>20</b>
2.1 Méthodes . . . . .	20
2.2 Echantillonnage . . . . .	26
2.3 État des lieux de la non-réponse . . . . .	28
2.4 Redressement . . . . .	29
2.5 Précision . . . . .	30
<b>3 Analyse des items</b>	<b>33</b>
3.1 Méthodologie . . . . .	33
3.2 Codage des réponses aux items . . . . .	36
3.3 Résultats . . . . .	40
<b>4 Modélisation</b>	<b>41</b>
4.1 Méthodologie . . . . .	41
4.2 Résultats . . . . .	48
4.3 Calcul des scores . . . . .	50
<b>5 Construction de l'échelle</b>	<b>50</b>
5.1 Méthode . . . . .	50
5.2 Caractérisation des groupes de niveaux . . . . .	52
5.3 Exemples d'items . . . . .	54
<b>6 Variables contextuelles et non cognitives</b>	<b>69</b>
6.1 Variables sociodémographiques et indice de position sociale . . . . .	69
6.2 Élaboration des questionnaires de contexte . . . . .	70
6.3 Motivation des élèves face à la situation d'évaluation . . . . .	70
<b>7 Annexe</b>	<b>73</b>
<b>Références</b>	<b>76</b>

## **Introduction**

La DEPP (Direction de l'évaluation, de la prospective et de la performance) conduit des évaluations de type CEDRE (Cycle des Évaluations Disciplinaires Réalisées sur Échantillon) depuis 2003.

Ce dispositif établit des bilans nationaux des acquis des élèves en fin d'école et en fin de collège. Reposant sur des épreuves standardisées, il couvre les compétences et les connaissances des élèves dans la plupart des domaines disciplinaires en référence aux programmes d'enseignement et au socle commun des connaissances et de compétences.

D'abord renouvelées tous les six ans, puis tous les cinq ans à partir de 2012, ces évaluations permettent d'assurer une comparabilité dans le temps. En 2015, CEDRE Compétences générales et langagières est remplacée par une nouvelle évaluation, intitulée CEDRE Compétences langagières et littératie. Le deuxième point de mesure, initialement prévu en 2020, a dû être reporté en 2021 en raison de la pandémie.

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.

La première partie de ce dossier décrit la méthodologie de l'enquête CEDRE Compétences langagières et littératie 2021, réalisée en fin de collège. La seconde partie s'intéresse à l'analyse des résultats.

# 1 Cadre d'évaluation

## 1.1 Objectifs

L'évaluation CEDRE Compétences langagières et littératie, réalisée en fin de collège, a pour objectif de faire le point sur les compétences des élèves relatives à leurs savoirs et leurs savoir-faire. Son renouvellement périodique permet une analyse de l'évolution des acquis des élèves dans le temps. Il s'agit ici de la deuxième prise d'informations. La précédente évaluation (2015) a été administrée essentiellement sur format papier, avec la passation, à titre exploratoire, d'un module numérique par un sous-échantillon. L'évaluation 2021 présente la particularité d'avoir été passée à la fois sous format papier et sous format numérique. Les items repris des évaluations antérieures ont été passés sur cahier. Tous les nouveaux items ont été conçus pour une passation sur ordinateur. L'ensemble des items de compréhension de l'écrit a servi à la constitution de l'échelle.

Les apports de cette évaluation portent sur :

- Les acquis des élèves en compréhension de l'écrit à l'issue du cycle 4 ;
- Les représentations de élèves concernant des stratégies de compréhension de l'écrit ou de l'oral.

L'éclairage apporté par l'enquête peut concerner tous les acteurs ayant un intérêt dans le système éducatif français : décideurs, enseignants sur le terrain, formateurs et chercheurs.

## 1.2 Connaissances et compétences visées

Le contenu de l'évaluation CEDRE Compétences Langagières et littératie se réfère au Programme du cycle 4 (BOEN spécial n°11 du 26 novembre 2015) concernant les attendus de fin de cycle en français dans les sous-domaines « Langage oral » et « Lecture et compréhension de l'écrit et de l'image » (Tableau 1). Parallèlement, conformément à la partie « Contributions essentielles des différents enseignements au socle commun » — qui stipule que « tous les champs disciplinaires concourent à la maîtrise de la langue »<sup>1</sup> (expression qui désigne les compétences langagières, dont la lecture et la compréhension de l'oral) — et en adéquation avec l'éducation aux médias et à l'information (EMI) — « présente dans tous les champs du savoir transmis aux élèves » et « prise en charge par tous les enseignements »<sup>2</sup> —, CEDRE CLL se rapproche des standards internationaux en dépassant le cloisonnement disciplinaire et en accordant une place non négligeable aux contenus relatifs aux médias et à la communication numérique (Tableau 2 et Tableau 3). L'objectif est ainsi d'évaluer les compé-

---

1. B.O. spécial n° 11 du 26 novembre 2015, p. 221.

2. *ibid.*, p. 382.

tences acquises tout au long de la scolarité dans le domaine de la « littératie »<sup>3</sup> (lettrisme) selon un cadre de compétences commun à la compréhension de l'écrit et de l'oral (Tableau 4).

---

3. « La littératie est la capacité d'une personne à mobiliser adéquatement, en contexte communicationnel synchrone ou asynchrone, les ressources et les compétences sémiotiques modales (ex : mode linguistique seul) et multimodales (ex : combinaison des modes linguistique, visuel et sonore) les plus appropriées à la situation et au support de communication (traditionnel et/ou numérique), à l'occasion de la réception (décryptage, compréhension, interprétation et évaluation) et/ou de la production (élaboration, création, diffusion) de tout type de message. » (Lacelle, Lebrun et Boutin, 2015)

TABLEAU 1 – Programme du cycle 4 – Français (B.O.E.N. spécial n° 11 du 26 novembre 2015)

<b>SOUS-DOMAINES</b>	<b>ATTENDUS DE FIN DE CYCLE</b>	<b>COMPÉTENCES ET CONNAISSANCES ASSOCIÉES</b>
Langage oral	Comprendre et interpréter des messages et des discours oraux complexes	Identification des visées d'un discours oral, hiérarchisation des informations qu'il contient, mémorisation des éléments importants
		Distinction de ce qui est explicite et de ce qui est sous-entendu dans un propos
Lecture et compréhension de l'écrit et de l'image	Lire des textes variés avec des objectifs divers	Adapter sa lecture à l'objectif poursuivi
		Adapter sa lecture au support et aux modes d'expression
		Reconnaître les implicites d'un texte et faire des inférences et hypothèses de lecture nécessaires
		Recourir à des stratégies de lecture diverses
	Lire des images, des documents composites (y compris numériques) et des textes non littéraires	Caractéristiques des différents documents étudiés (scientifiques, médiatiques, composites...)
	Lire et comprendre des images fixes ou mobiles variées empruntées à la peinture, aux arts plastiques, à la photographie, à la publicité et au cinéma en fondant sa lecture sur que les outils d'analyse simples. Situer les œuvres dans leur contexte historique et culturel.	Éléments d'analyse de l'image
		Relations textes littéraires, images illustratives et adaptations cinématographiques
	Lire des œuvres littéraires et fréquenter des œuvres d'art	Différents genres littéraires
		Caractéristiques et enjeux de l'environnement médiatique et numérique
		Les œuvres majeures depuis l'Antiquité, leur contexte et le dialogue entre les arts
Élaborer une interprétation de textes littéraires	Formuler des impressions de lecture	
	Percevoir un effet esthétique et en analyser les sources	
	Situer une œuvre dans son contexte pour éclairer ou enrichir sa lecture et établir des relations entre des œuvres littéraires et artistiques	

Ce tableau restitue exclusivement les attendus en compréhension de l'écrit et de l'oral. Les éléments évalués dans le CEDRE CLL paraissent en gras.

TABLEAU 2 – Programme du cycle 4 – Éducation aux médias et à l'information  
(B.O.E.N. spécial n° 11 du 26 novembre 2015)

COMPÉTENCES TRAVAILLÉES	DOMAINES DU SOCLE
Utiliser les médias et les informations de manière autonome	2
<b>Exploiter l'information de manière raisonnée :</b> <ul style="list-style-type: none"> <li>— Distinguer les sources d'information, s'interroger sur la validité et sur la fiabilité d'une information, son degré de pertinence</li> <li>— S'entraîner à distinguer une information scientifique vulgarisée d'une information pseudo-scientifique grâce à des indices textuels ou paratextuels et à la validation de la source</li> <li>— Apprendre à distinguer subjectivité et objectivité dans l'étude d'un objet médiatique</li> <li>— Découvrir les représentations du monde véhiculées par les médias</li> <li>— S'interroger sur l'influence des médias sur la consommation et la vie démocratique</li> </ul>	1,3,5
Utiliser les médias de manière responsable	3
Produire, communiquer, partager des informations	1

Ce tableau ne détaille les compétences travaillées que pour la thématique « Exploiter l'information de manière raisonnée », en lien avec l'aspect compréhension du domaine 1 du socle. Les éléments évalués dans CEDRE CLL paraissent en gras.

TABLEAU 3 – Typologie des supports exploités en compréhension de l'écrit et de l'oral dans CEDRE CLL

DOMAINE	PASSATION	SUPPORTS
<b>Compréhension de l'écrit</b>	<b>Passation sur papier</b>	Texte littéraire
		Texte documentaire continu
		Document composite de source unique
	<b>Passation sur ordinateur</b>	Document composite de sources multiples
Document « numérique » (messagerie, chat, blog, forum)		
Oral préparé : reportage, chronique, etc.		
<b>Compréhension de l'oral (audio ou vidéo)</b>	<b>Passation sur ordinateur</b>	Oral non préparé (expert) : débat, interview, etc.)
		Oral non-préparé (non expert) : micro-trottoir, antenne libre, etc.)

TABLEAU 4 – Cadre de compétences en compréhension de l'écrit et de l'oral de CEDRE CLL

COMPÉTENCES	ASPECTS
<b>Retrouver une information explicite (localiser/mémoriser)</b>	<ul style="list-style-type: none"> <li>— Retrouver une information explicite immédiatement repérable</li> <li>— Retrouver une information explicite non immédiatement repérable</li> </ul>
<b>Mettre en relation des informations et réaliser des inférences (inférer/intégrer)</b>	<ul style="list-style-type: none"> <li>— Réaliser une inférence locale</li> <li>— Réaliser une inférence de connaissance ou « pragmatique »</li> <li>— Mettre en relation et intégrer des informations</li> </ul>
<b>Rendre compte du sens global et identifier la visée (résumer/organiser)</b>	<ul style="list-style-type: none"> <li>— Organiser les idées principales d'un document</li> <li>— Dégager les idées principales d'un document</li> <li>— Identifier la visée</li> </ul>
<b>Évaluer l'information et faire preuve d'esprit critique (réfléchir/évaluer)</b>	<ul style="list-style-type: none"> <li>— Exprimer son opinion sur le document en convoquant ses connaissances ou son expérience</li> <li>— Évaluer la pertinence de l'information et les moyens utilisés pour l'exprimer</li> </ul>

### 1.3 Construction du test

Le cycle complet de cette évaluation s'est déroulé sur six années au lieu de cinq en raison de la pandémie de COVID19.

TABLEAU 5 – Déroulé de l'évaluation Maitrise de la langue 2021

Année 1 (2017-2018) : Création d'items	<ul style="list-style-type: none"> <li>— Définition du cadre de l'évaluation à partir des programmes avec l'Inspection Générale de l'Éducation du Sport et de la Recherche (IGESR).</li> <li>— Conception d'items au format numérique prenant en compte le cadre de l'évaluation.</li> </ul>
Année 2 (2018-2019) : Création d'items, étude de comparabilité et expérimentation de l'évaluation numérique	<ul style="list-style-type: none"> <li>— Etude de comparabilité (Bridge Study) entre le format papier et le format numérique. mars-avril 2019.</li> <li>— Construction de l'expérimentation à partir des items créés.</li> <li>— Passation de l'expérimentation numérique en mai 2019 auprès d'un échantillon représentatif d'élèves.</li> </ul>
Année 3 (2019-2020) : Évaluation	<ul style="list-style-type: none"> <li>— Analyse des résultats de l'expérimentation.</li> <li>— Construction de l'évaluation numérique à partir d'une sélection d'items validés après l'expérimentation.</li> <li>— Passation de l'évaluation prévue en mai 2020 auprès d'un échantillon représentatif d'élèves : annulée en raison de la pandémie.</li> </ul>
Année 4 (2020-2021) : Évaluation reportée	<ul style="list-style-type: none"> <li>— Passation de l'évaluation en mai 2021 auprès d'un échantillon représentatif d'élèves.</li> </ul>
Année 5 (2021-2023) : Analyse des résultats et publications	<ul style="list-style-type: none"> <li>— Analyse des résultats de l'évaluation.</li> <li>— Publication d'une Note d'Information et d'un article dans le RERS.</li> <li>— Préparation du dossier.</li> </ul>

#### 1.3.1 Conceptions des items

Les items produits dans une évaluation CEDRE sont le fruit d'un travail collectif des concepteurs, encadrés par un chargé d'étude, personnel du Bureau de la

conception et du pilotage des évaluations des élèves (DEPP B2-1), sous la responsabilité du chef de bureau. Les concepteurs sont des professeurs de Lettres enseignant en collège ou lycée dans des académies diverses. Des experts issus de l'université, de la formation ou de l'Inspection générale participent à certaines réunions.

Les concepteurs recherchent des supports selon le cahier des charges qui leur est fourni et proposent des items aux différents formats de question demandés. Ces items font l'objet d'une relecture au sein du groupe et sont soumis à un « cobayage » dans les classes des concepteurs. Après la finalisation des items, le chargé d'étude les transmet à l'Inspection générale pour leur validation pédagogique. Une expérimentation organisée auprès d'un échantillon national dans des conditions réelles de passation permet de valider les items sur le plan statistique. Les items retenus peuvent être intégrés à l'évaluation.

Une application ad hoc (GEODE) est utilisée en interne pour faciliter la production des items, les standardiser, les éditer, les stocker et gérer l'évaluation dans son ensemble.

### 1.3.2 Constitution du test

Le test est constitué d'un ensemble d'unités ou situations (ensemble d'items composant un exercice cohérent), organisées en « blocs », eux-mêmes répartis dans des cahiers pour la séquence passée sur papier ou dans des modules pour la séquence passée sur ordinateur.

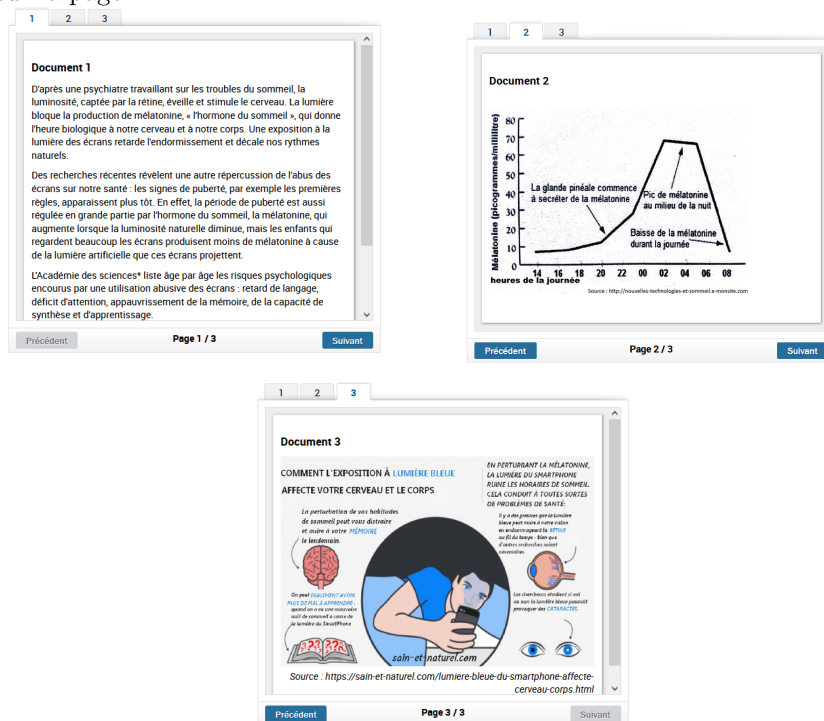
La séquence papier CEDRE Compétences langagières et littératie 2021 est intégralement composée d'items repris à l'identique par rapport aux évaluations antérieures : ce sont les « items d'ancrage ». Cette séquence est constituée de 21 unités réparties dans 11 cahiers. Toutes concernent la compréhension de l'écrit. L'évaluation numérique est entièrement constituée d'items nouveaux qui ont fait l'objet d'une expérimentation en 2019. Les items ont été créés pour une passation sur ordinateur. Les élèves doivent également être équipés d'un casque audio pour les exercices de compréhension de l'oral.

La séquence numérique CEDRE Compétences Langagières et Littératie 2021 comporte au total 208 items retenus sur la base de leur validité statistique après expérimentation en 2019. Ils appartiennent à 24 situations réparties dans 11 modules. Ces unités concernent 2 domaines :

- La compréhension de l'écrit (71 items / 14 unités)
- La compréhension de l'oral (32 items / 10 unités)

En numérique, les supports de compréhension de l'écrit sont présentés dans un format appelé « liseuse ». La liseuse, présente sur la gauche en regard de chaque question, permet de naviguer entre différents documents ou différentes parties

FIGURE 1 – Exemple de liseuse à 3 onglets avec barre de défilement à utiliser sur la page 1.



du support à l'aide d'onglets. Dans certains cas, l'élève peut avoir à utiliser la barre de défilement pour consulter l'intégralité du support. La lecture sur liseuse peut avoir un impact sur les stratégies et la mémoire de lecture par rapport à la lecture traditionnelle sur papier.

La compréhension orale est évaluée à partir de 11 supports faisant varier les types d'oral (lu/non lu), les types d'interaction (monologue, échange), le type de texte (explicatif, argumentatif, narratif), la situation par rapport à l'élève (proche ou éloignée) et enfin la position du locuteur (expert ou non expert). Deux de ces supports sont des extraits vidéo ; les autres sont des supports audio. Quant à la compréhension de l'écrit, les 36 unités qui la composent portent sur des textes littéraires ou documentaires continus, des documents composites (source unique), des groupements thématiques (sources multiples) et des documents relatifs à la communication numérique (chats, blogs, forums ou courriels fictifs).

Pour mémoire, la compréhension est évaluée à travers 4 compétences :

- Retrouver une information explicite (localiser/mémoriser) ;

FIGURE 2 – Répartition des items selon l'année de conception et le domaine

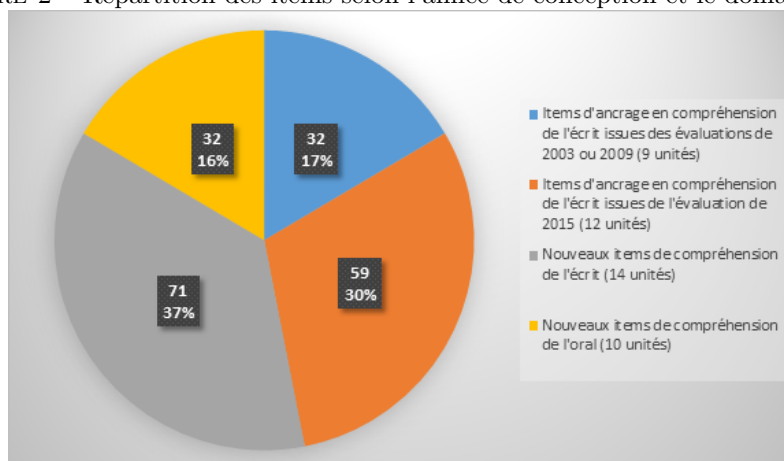
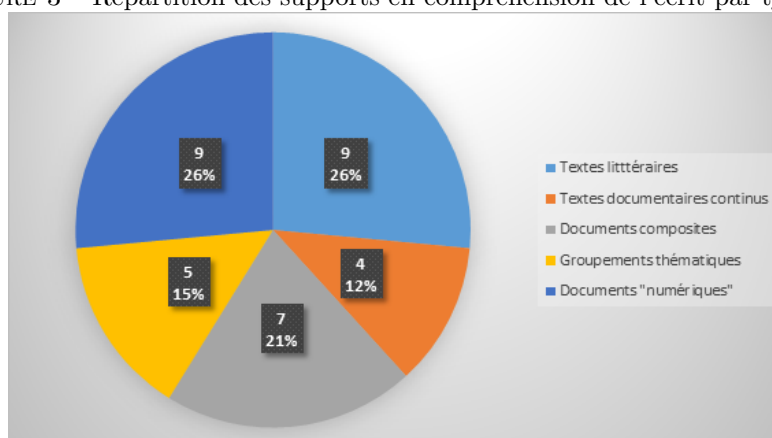


FIGURE 3 – Répartition des supports en compréhension de l'écrit par type



- Mettre en relation des informations et réaliser des inférences (inférer/intégrer) ;
- Rendre compte du sens global et identifier la visée (résumer/organiser) ;
- Évaluer l'information et faire preuve d'esprit critique (réfléchir/évaluer).

Selon s'il s'agit de compréhension de l'écrit ou de compréhension de l'oral, la répartition des compétences diffère. La compréhension de l'écrit, en cohérence avec les programmes, interroge particulièrement les compétences d'inférence et d'interprétation, tandis que la compréhension de l'oral, notamment en raison des modalités de passation — écoutes limitées à deux fois avant l'accès aux questions — privilégie la compréhension du sens global ou de la visée.

FIGURE 4 – Répartition des items en compréhension de l'écrit par compétence

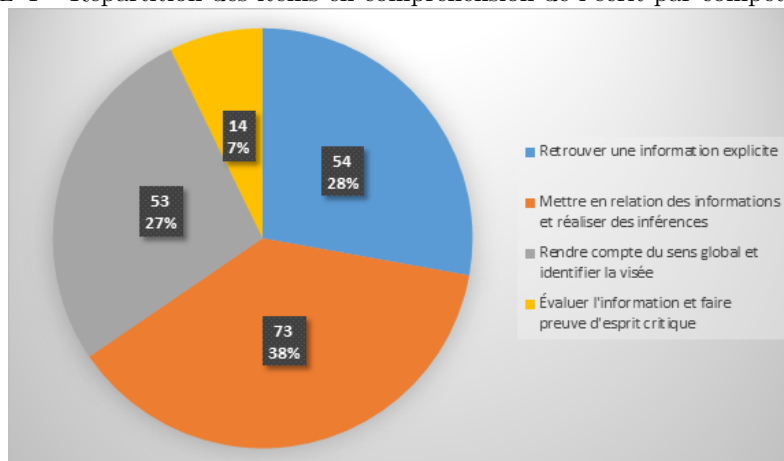
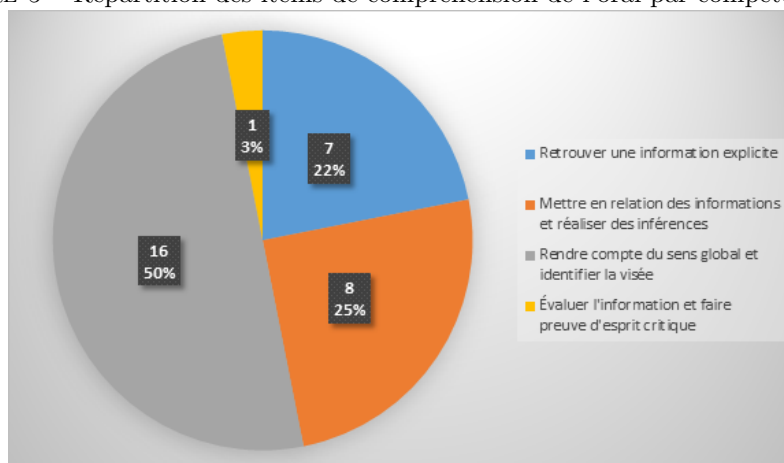


FIGURE 5 – Répartition des items de compréhension de l'oral par compétence



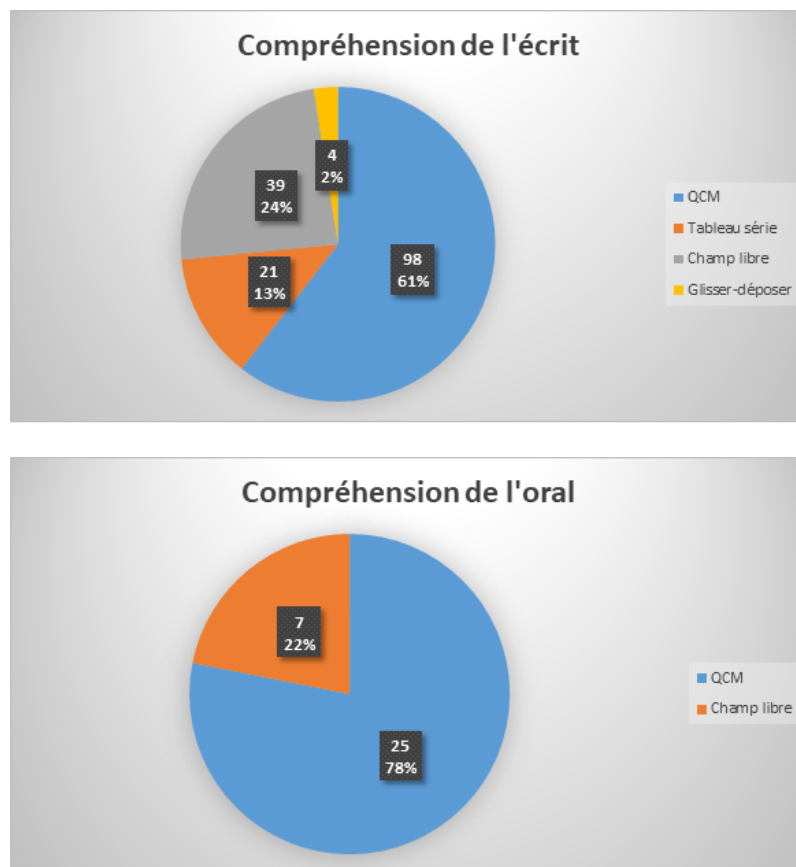
### 1.3.3 Formats d'items

Trois formats de questions ont été principalement utilisés pour l'évaluation papier : QCM, tableau série, champ libre. Quelques items de la séquence numérique exploitent un format de type glisser-déposer permettant d'établir un classement ou une mise en ordre.

#### 1.3.3.a Le questionnaire à choix multiples (QCM)

Le QCM (questionnaire à choix multiples) permet d'investiguer un large champ de connaissances et de compétences dans un temps restreint, et de ne pas conditionner la réussite de l'item à la subjectivité du correcteur. Les réponses choisies par les élèves ne sont pas ambiguës, puisqu'elles résultent d'un choix unique. Les

FIGURE 6 – Répartition des items selon leur format en compréhension de l'écrit et de l'oral



compétences rédactionnelles de l'élève ne sont pas mises en jeu dans la validation de sa réponse. En multipliant les questions relatives à une même compétence, on fiabilise la prise d'information. La question à choix multiples est constituée de deux composantes :

- L'amorce, qui pose la question ou définit le problème. L'amorce doit être simple et facilement compréhensible ;
- Les 4 propositions, parmi lesquelles figure une seule réponse correcte ; les trois autres propositions sont appelées « distracteurs ». En passation numérique, les propositions sont mélangées aléatoirement.

FIGURE 7 – Exemple de format QCM

**ÉCRANS ET SANTÉ**

1 2 3

**Document 1**

D'après un psychiatre travaillant sur les troubles du sommeil, la luminosité, captée par la rétine, éveille et stimule le cerveau. La lumière bloque la production de mélatonine, « l'hormone du sommeil », qui donne l'heure biologique à notre cerveau et à notre corps. Une exposition à la lumière des écrans retarde l'endormissement et décale nos rythmes naturels.

Des recherches récentes révèlent une autre répercussion de l'abus des écrans sur notre santé : les signes de puberté, par exemple les premières règles, apparaissent plus tôt. En effet, la période de puberté est aussi régulée en grande partie par l'hormone du sommeil, la mélatonine, qui augmente lorsque la luminosité naturelle diminue, mais les enfants qui regardent beaucoup les écrans produisent moins de mélatonine à cause de la lumière artificielle que ces écrans projettent.

L'Académie des sciences\* liste âge par âge les risques psychologiques encourus par une utilisation abusive des écrans : retard de langage, déficit d'attention, appauvrissement de la mémoire, de la capacité de synthèse et d'apprentissage.

Précédent Page 1 / 3 Suivant

**Question 6/7**

**L'un des dangers des écrans ci-dessous est évoqué dans le document 3 mais pas dans le document 1 : lequel ?**

- les troubles du sommeil
- les problèmes de mémoire
- les dégâts sur la rétine
- les difficultés à apprendre

### 1.3.3.b Le tableau série

Le format tableau série permet de varier les processus cognitifs attendus des élèves et parfois d'interroger un niveau plus détaillé de compréhension. Le score n'est pas calculé sur une seule réponse mais sur l'ensemble des propositions. Pour chaque tableau, un seuil de réponses correctes à partir duquel un crédit complet est accordé est défini. Ce seuil peut exiger une réponse correcte à chaque ligne. En passation numérique, les lignes du tableau sont mélangées aléatoirement.

FIGURE 8 – Exemple de format tableau série

ÉCRANS ET SANTÉ

1
2
3

**Document 3**

Comment l'exposition à la lumière bleue affecte votre cerveau et le corps. En perturbant la mélatonine, la lumière du smartphone ruine les horaires de sommeil. Cela conduit à toutes sortes de problèmes de Santé : • Il y a des preuves que la lumière bleue peut nuire à notre vision en endommageant la rétine au fil du temps bien que d'autres recherche soient nécessaires. Les chercheurs étudient si ou non la lumière bleue pourrait provoquer des cataractes. • La perturbation de vos habitudes de sommeil peut vous distraire et nuire à votre mémoire le lendemain • On peut également avoir plus de mal à apprendre quand on a eu une mauvaise nuit de sommeil à cause de la lumière du smartphone L'image montre que regarder son téléphone avant de dormir fait partie des conduites à risque . Source : <https://sain-et-naturel.com/lumiere-bleue-du-smartphone-affecte-cerveau-corps.html>

Source : <https://sain-et-naturel.com/lumiere-bleue-du-smartphone-affecte-cerveau-corps.html>

Précédent
Page 3 / 3
Suivant

**Question 5/7**

**Les affirmations ci-dessous concernant le document 3 sont-elles vraies ou fausses ?**

	Vrai	Faux
La couleur bleue a été choisie pour faire moins peur au lecteur.	<input type="checkbox"/>	<input type="checkbox"/>
Les mots en bleu rappellent le thème principal du document 3.	<input type="checkbox"/>	<input type="checkbox"/>
Les mots en bleu soulignent les idées essentielles à retenir.	<input type="checkbox"/>	<input type="checkbox"/>
Les mots sont en bleu car c'est la couleur des yeux du personnage.	<input type="checkbox"/>	<input type="checkbox"/>
La couleur bleue doit être la couleur préférée de l'auteur du document 3.	<input type="checkbox"/>	<input type="checkbox"/>

### 1.3.3.c Le champ libre

Le champ libre est un format d'item dans lequel figure un espace où l'élève peut produire sa réponse. Celle-ci peut prendre la forme d'un texte, d'une phrase, d'une lettre, d'un chiffre ou d'un nombre. Le codage de ces questions est pour leur majorité l'objet d'une correction experte effectuée par un collège d'enseignants selon un guide de codage précis. De par son caractère ouvert, le champ libre est entre autres particulièrement utile pour évaluer les compétences d'esprit critique des élèves.

FIGURE 9 – Exemple de format champ libre

**ÉCRANS ET SANTÉ**

1 2 3

**Document 1**

D'après une psychiatre travaillant sur les troubles du sommeil, la luminosité, captée par la rétine, éveille et stimule le cerveau. La lumière bloque la production de mélatonine, « l'hormone du sommeil », qui donne l'heure biologique à notre cerveau et à notre corps. Une exposition à la lumière des écrans retarde l'endormissement et décale nos rythmes naturels.

Des recherches récentes révèlent une autre répercussion de l'abus des écrans sur notre santé : les signes de puberté, par exemple les premières règles, apparaissent plus tôt. En effet, la période de puberté est aussi régulée en grande partie par l'hormone du sommeil, la mélatonine, qui augmente lorsque la luminosité naturelle diminue, mais les enfants qui regardent beaucoup les écrans produisent moins de mélatonine à cause de la lumière artificielle que ces écrans projettent.

L'Académie des sciences\* liste âge par âge les risques psychologiques encourus par une utilisation abusive des écrans : retard de langage, déficit d'attention, appauvrissement de la mémoire, de la capacité de synthèse et d'apprentissage.

Précédent Page 1 / 3 Suivant

**Question 7/7**

**Quel est le document qui inspire le plus confiance entre le document 1 et le document 3 ?  
Expliquez la raison de votre choix.**

#### 1.3.3.d Le glisser-déposer

Format exclusivement réservé à la passation numérique, le glisser-déposer permet notamment l'évaluation des compétences d'ordonnancement ou de hiérarchisation des élèves.

FIGURE 10 – Exemple de format glisser-déposer

**APPARTEMENT**

Loubna Khacer vit seule à Libourne avec sa fille de cinq ans. Elle vient de décrocher un emploi au centre commercial de Mériadeck à Bordeaux. Ne possédant pas de voiture, Loubna souhaite se rapprocher de son lieu de travail. Elle contacte Thierry Demestre qui lui propose plusieurs appartements à la location.

1 2

De : Loubna Khacer < LoubnaKhacer@tospeed.fr > Envoyé : Mer 25/07/2017 9:05 À : Thierry Demestre < T.Demestre@tospeed.fr > Objet : Demande de location Monsieur, J'ai appris par Christophe Lebas, votre neveu, que vous proposiez plusieurs appartements à la location à Bordeaux. Je suis à la recherche d'un logement de préférence situé dans le secteur du centre commercial. Je souhaiterais retrouver un appartement disposant de deux chambres, situé de préférence au rez-de-chaussée ou accessible par ascenseur, pour un loyer maximal de 800 euros charges comprises. J'aimerais savoir si vous aviez encore des biens de ce type disponibles et si vous pouviez, dans ce cas, me transmettre un descriptif incluant le montant du loyer. Je dois débiter mon nouvel emploi le 15 septembre prochain et aimerais donc déménager au plus tard le 25 Août. Cordialement.

**Question 2/4**

**Remettez dans l'ordre les différentes étapes de la demande de Loubna (page 1).**

- détail des attentes
- secteur souhaité
- délais à respecter
- demande d'informations

Précédent Page 1 / 2 Suivant

## 1.4 Présentation générale de la passation

L'évaluation s'est déroulée du lundi 17 mai au vendredi 28 mai 2021. Des consignes définies au niveau national ainsi que la durée des exercices étaient données dans un guide de passation afin que les conditions de recueil de données et de passation soient rigoureusement les mêmes dans tous les collèges concernés. La passation s'organisait en deux séquences de 45 minutes effectives de travail : une séquence sur cahier en premier lieu et une séquence sur ordinateur dans un deuxième temps.

### 1.4.0.a L'épreuve papier

Elle était composée de 21 situations de compréhension de l'écrit réparties dans 11 cahiers différents. Dans une même classe, les élèves n'avaient pas tous le même cahier. Cependant, des situations identiques pouvaient se retrouver d'un cahier à l'autre.

### 1.4.0.b L'épreuve numérique

Pour la passation numérique, 24 situations de compréhension de l'écrit ou de l'oral étaient réparties dans 11 modules différents. Des situations identiques pouvaient se retrouver d'un module à l'autre. À l'issue de l'épreuve numérique, un questionnaire de contexte composé de 24 questions était soumis aux élèves.

## 1.4.1 Le codage des items

Deux types d'items sont à coder dans les évaluations CEDRE :

- Les questions fermées corrigées de façon automatisée (après numérisation dans le cas du format papier) ;
- Les questions ouvertes appelant une réponse rédigée. Une correction experte est alors nécessaire : des enseignants du second degré sont recrutés pour cette mission.

Pour les questions ouvertes, un guide de codage décrit les critères de réussite et fournit des exemples de réponses correspondant aux différents codes de réponse correcte ou incorrecte. Une formation des correcteurs à ce guide de codage, avec une série de réponses d'entraînement pour chaque item à correction experte, permet d'harmoniser les codages et de limiter la subjectivité dans la correction. Un système de double correction appliqué sur 10 % des items permet de vérifier la fiabilité du codage. Le codage des réponses aux items d'ancrage, repris d'un cycle d'évaluation à un autre, est strictement identique à celui des années précédentes.

## 2 Sondage

### 2.1 Méthodes

#### 2.1.1 Tirage équilibré de classes de 3e

De manière générale, pour le secondaire, deux options de tirage peuvent être considérées : soit un sondage par grappe en sélectionnant un échantillon de classes et tous les élèves des classes tirées au sort participent à l'évaluation ; soit un premier degré qui concerne les établissements puis un second degré où un nombre d'élèves fixe dans chaque établissement est sélectionné<sup>4</sup>. Les évaluations CEDRE suivent la première option tandis que l'évaluation PISA suit la seconde. Des simulations ont permis de montrer que les niveaux de précision des deux options sont très proches, dès lors que le tirage est équilibré (cf. encadré « Tirage d'établissement *versus* tirage de classes »). Le choix de sondages par grappe est motivé par la facilité de gestion. En effet, le fait de sélectionner tous les élèves d'une classe au collège permet d'éviter de mettre en place des procédures de tirage au sort d'élèves une fois les établissements tirés ; cela facilite également l'organisation de l'évaluation pour le personnel de l'établissement.

On note  $U$  la population visée par une évaluation donnée,  $Y$  la variable d'intérêt (typiquement le score à l'évaluation, ou bien une indicatrice de difficulté),  $X$  une variable auxiliaire, c'est-à-dire connue pour l'ensemble des élèves de la population  $U$ . Un échantillon  $S$  d'élèves est sélectionné dans la population  $U$ . Chaque élève  $i$  a la probabilité  $\pi_i$  d'être sélectionné dans l'échantillon  $S$  (probabilité d'inclusion). Enfin, les poids de sondages, définis comme les inverses des probabilités d'inclusion  $\pi_i$ , sont notés  $d_i$ .

Un échantillon équilibré est un échantillon qui est représentatif de la population au regard de certaines variables auxiliaires. Cela signifie que dans un échantillon équilibré, l'estimateur du total d'une variable auxiliaire  $X$  sera exactement égal au vrai total de la variable  $X$  dans la population.

Cette propriété s'écrit :

$$\sum_{i \in S} \frac{X_i}{\pi_i} = \sum_{i \in U} X_i \quad (1)$$

---

4. Dans ce second cas, les établissements sont tirés proportionnellement à leur taille (nombre d'élèves). En effet, une fois que les établissements sont échantillonnés, un nombre fixe d'élèves est alors sélectionné quel que soit l'établissement. Par conséquent, les élèves des grands établissements ont moins de chance d'être tirés au sort que les élèves des petits établissements. Le tirage proportionnel à la taille permet ainsi de rétablir l'égalité des probabilités de tirage.

### **Tirage d'établissements *versus* Tirage de classes**

Pour faciliter la logistique dans les collèges, nous réalisons un tirage de classes de 3e, puis tous les élèves de la classe sélectionnée passent l'évaluation. On peut donc s'interroger sur la perte de la précision liée à cet effet de grappe.

Pour comparer la précision entre un tirage d'établissement et un tirage de classes, nous avons réalisé des simulations à partir de la base des notes au brevet en 2009 (Garcia, Le Cam, & Rocher, 2015).

Nous avons comparé deux stratégies d'échantillonnage. Il s'agit à chaque fois d'échantillons stratifiés à deux degrés :

- Tirage équilibré d'établissement puis tirage de 30 élèves dans chaque établissement sélectionné ;
- Tirage équilibré de classe puis sélection de tous les élèves des classes sélectionnées.

La stratification a été effectuée selon le secteur d'enseignement et dans chaque strate 2 000 élèves ont été échantillonnés.

Pour chacune des deux stratégies, 1 000 échantillons ont été tirés. Puis on calcule la moyenne des erreurs standards des notes moyennes en français, mathématiques et histoire-géographie. Le tableau ci-dessous montre que les deux stratégies de tirage ont des niveaux équivalents de précision.

#### **Comparaison des erreurs standards (Garcia et al., 2015)**

	Echantillon équilibré d'établissements	Echantillon équilibré de classes
Français	0,07	0,07
Mathématiques	0,11	0,11
Histoire-Géographie	0,08	0,08

Les échantillons équilibrés ont donc comme propriété de fournir une photographie parfaite de la population, au regard des variables auxiliaires connues, ce que ne garantit pas une procédure aléatoire simple d'échantillonnage. En théorie, ils permettent également d'améliorer la précision des estimateurs s'il existe un lien entre la variable d'intérêt et les variables auxiliaires.

Le tirage équilibré est réalisé grâce à la fonction `samplecube` du package `sampling` sur R. La documentation complète est disponible sur le site Internet de l'INSEE (Rousseau & Tardieu, 2004). L'algorithme permet de choisir de manière aléatoire un échantillon parmi tous les échantillons possibles respectant les

contraintes reposant sur les variables auxiliaires. Il se déroule en deux phases : une « phase de vol » et une « phase d'atterrissage ». Durant la phase de vol, toutes les contraintes sont respectées. Elle se termine si un échantillon équilibré de manière parfaite est trouvé ou s'il n'est pas possible de trouver un échantillon en respectant toutes les contraintes. Si la phase de vol n'a pas abouti à un échantillon, la phase d'atterrissage débute. Elle consiste au relâchement des contraintes et au choix optimal de l'échantillon selon le critère choisi par l'utilisateur (ordre de priorité sur les contraintes, relâchement de la contrainte avec un coût minimal sur l'équilibrage ou garantie d'un échantillon de taille fixe).

Par ailleurs, au moment du tirage de l'échantillon, les collèves dont une classe a déjà été sélectionnée pour une autre évaluation la même année sont exclus 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 équilibré 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. 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 :

$$\forall k = 1 \dots K, \sum_{i \in S'} w_i X_i^k = \sum_{i \in U} X_i^k \quad (2)$$

Ils sont obtenus par minimisation de l'expression  $\sum_{i \in S'} d_i G\left(\frac{w_i}{d_i}\right)$  où  $G$  désigne une fonction de distance, sous les contraintes définies dans l'équation 2.

### Tirage équilibré 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 équilibré 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) est en cours de développement avec une application à des données issues d'évaluations standardisées (Christine & Rocher, 2012).

#### Formulation du problème et notations

Un échantillon  $S_1$  a été tiré. Il est connu et les probabilités d'inclusion des établissements  $\pi_j^1$  sont également connues. On souhaite alors tirer un échantillon  $S_2$  dans la population  $U$  avec les probabilités  $\pi_j^2$ , 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  $\pi_j^{2/S_1}$  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  $\lambda_j = \frac{\pi_j^2}{1 - \pi_j^1}$

#### Équilibrage

On souhaite maintenant que l'échantillon  $S_2$  soit équilibré selon certaines

variables (nombre d'élèves en retard, etc.). Soit  $X$  une variable d'équilibre, la condition s'écrit :

$$\sum_{j \in S_2} \frac{X_j}{\pi_j^2} = \sum_{j \in U} X_j$$

Pour arriver à ce résultat, le principe est de tirer  $S_2$  dans  $U(S_1)$  avec les probabilités d'inclusion  $\lambda_j$  et avec une condition d'équilibre sur la variable  $X_j/(1 - \pi_j^1)$ .

Ainsi, on aura :

$$\sum_{j \in S_2} \frac{X_j}{\pi_j^2} = \sum_{j \in S_2} \frac{X_j}{\lambda_j(1 - \pi_j^1)} = \sum_{j \in U(S_1)} \frac{X_j}{1 - \pi_j^1}$$

Or, en espérance on a

$$E\left(\sum_{j \in U(S_1)} \frac{X_j}{1 - \pi_j^1}\right) = E\left(\sum_{j \in U} \frac{X_j}{1 - \pi_j^1} I_{j \notin S_1}\right) = \sum_{j \in U} X_j$$

La condition d'équilibre initiale est donc remplie.

### 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 réplification, 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 ( $w_i$ ) 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 évaluation CEDRE au collège est celui des élèves de 3e générale scolarisés dans des collèges publics et privés sous contrat de France métropolitaine.

La base de sondage utilisée est la base dite Scolarité construite par la DEPP. C'est une base de données individuelles anonymes contenant de nombreuses informations sur les élèves scolarisés une année scolaire donnée (date de naissance, PCS des parents, etc.). Nous disposons également d'informations sur les établissements scolaires, comme par exemple le secteur d'enseignement. Ces informations, qualifiées de variables auxiliaires, peuvent être utilisées au moment du tirage des échantillons, pour définir les variables de stratification. Préalablement au tirage, les établissements des échantillons d'autres opérations d'évaluations de la DEPP sont retirés de la base de sondage.

### Stratification

Une stratification est réalisée en fonction du secteur d'enseignement :

1. Public hors éducation Prioritaire (PU)
2. Public en éducation prioritaire (EP)
3. Privé (PR)

### Modalités de sélection

Le tirage est à deux degrés. Le premier degré de sondage est composé de classes (et non de collèges) tirées dans chaque strate avec allocation proportionnelle. Le deuxième degré de sondage consiste à interroger tous les élèves de la classe sélectionnée (tirage par grappe). La fonction `samplecube` du package `sampling` est utilisée pour garantir des échantillons équilibrés sur la base de sondage selon certaines variables

Dans chacune des 3 strates, le tirage est équilibré sur les variables suivantes :

- Le nombre total d'élèves de 3e
- L'indice de position sociale (Rocher, 2016)
- Le nombre d'élèves de 3e en retard dans la population
- Le nombre de garçons de 3e dans la population

### Echantillon 2021

L'échantillon vise 8 000 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 - Compétences langagières et littératie Collège

	Établissements	Elèves
Etab. accueillant des élèves de 3e	8 549	864 452
On retire les COM	8 509	860 129
On retire les étab hors contrat	8 228	856 590
On retire les EREA	8 163	855 428
On retire les UPE2A	8 141	854 082
On retire les ULIS	8 136	852 854
On ne garde que les collèges	6 940	825 291
On ne garde que les 3e générales	6 936	803 688
On retire Mayotte	6 914	796 646
On exclut les établissements déjà sélectionnés	5 309	593 261
Base CEDRE 3e	5 309	593 261

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

TABLEAU 7 – Répartition dans la base de sondage - Compétences langagières et littératie Collège

Strate	Établissements	Élèves
1.PU	4 197	499 144
2.REP	1 069	125 789
3.Privé	1 648	171 704
<b>Total</b>	<b>6 914</b>	<b>796 637</b>

**Échantillon**

Le tableau 8 présente la répartition de l'échantillon selon le secteur d'enseignement. Au total, 302 établissements ont été sélectionnés.

TABLEAU 8 – Répartition dans l'échantillon - Compétences langagières et littératie Collège

Strate	Établissements	Élèves
1. PU hors EP	111	3 006
2. EP	120	3 004
3. Privé	71	2 014
<b>Total</b>	<b>302</b>	<b>8 024</b>

## 2.3 État des lieux de la non-réponse

### 2.3.1 Non-réponse totale

Parmi la non-réponse totale, nous distinguons la non-réponse des établissements de la non-réponse des élèves des établissements participants. Les chiffres suivants ont été observés pour 2021.

85.8 % des établissements de l'échantillon ont répondu à l'évaluation (tableau 9).

83.9 % des effectifs attendus ont participé (tableau 10).

TABLEAU 9 – Non-réponse des établissements - Compétences langagières et littératie Collège

Strate	Nb établissements attendus	Nb établissements répondants	% d'établissements répondants
1.PU	111	97	87.4 %
2.REP	120	110	91.7 %
3. Privé	71	52	73.2%
<b>Total</b>	<b>302</b>	<b>259</b>	<b>85.8 %</b>

TABLEAU 10 – Non-réponse des élèves - Compétences langagières et littératie Collège

Strate	Nb élèves attendus	Nb élèves répondants	% d'élèves répondants
1. PU hors EP	3 006	2 553	84.9 %
2. EP	3 004	2 722	90.6 %
3. Privé	2 014	1 1459	72.4 %
<b>Total</b>	<b>8 024</b>	<b>6 734</b>	<b>83.9 %</b>

### 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. En 2021 cela concernait 1.3 % des élèves.

### 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 module.
- La non-réponse terminale : un élève s'est arrêté avant la fin du module 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 les dernières réponses d'un élève sont vides alors il n'a pas eu le temps de les voir. La non réponse terminale a été étudiée par séquence et par module. Toutes les observations en non-réponse terminale sont enlevées des données.

La non-réponse terminale représente en moyenne :

- 3.3 % des élèves en situation de NRT sur la séquence papier. Parmi eux la NRT représente en moyenne 4.5 items
- 2.8 % des élèves en situation de NRT sur la séquence numérique. Parmi eux la NRT représente en moyenne 2.6 items

## 2.4 Redressement

La correction de la non-réponse totale est effectuée en deux étapes. D'abord une post-stratification, selon les strates de sondage :

- 1. Public hors EP
- 2. REP
- 3. Privé sous contrat

Ensuite un calage sur marge est effectué en utilisant les deux variables de calage suivantes :

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

TABLEAU 11 – Comparaison entre les marges de l'échantillon et les marges dans la population - Compétences langagières et littératie Collège

Variable	Modalité	Marge pré-calage sur marges	Marge post-calage	Marge population	Pourcentage pré-calage	Pourcentage post-calage	Pourcentage population
Retard	1	69 968	90 220	90 220	9,0	11,6	11,6
	2	704 549	687 015	687 015	91,0	88,4	88,4
Sexe	1	381 664	389 459	389 459	49,3	50,1	50,1
	2	392 852	387 776	387 776	50,7	49,9	49,9
Strate	1	487 716	487 716	487 716	62,7	62,7	62,7
	2	122 899	122 899	122 899	15,8	15,8	15,8
	3	166 620	166 620	166 620	21,4	21,4	21,4

## 2.5 Précision

L'erreur standard de la moyenne (SEM) est calculée comme l'écart type divisé par la racine carrée de la taille de l'échantillon :

$$SEM = \frac{\sigma}{\sqrt{n}} \quad (3)$$

où  $\sigma$  est l'écart type de la population et  $n$  est la taille de l'échantillon (tableau 12).

TABLEAU 12 – Scores moyens et erreurs standard associées - Compétences langagières et littératie Collège

Année	Score moyen	Erreur standard
2015	250	1.21
2021	247	1.25

Pour savoir par exemple si l'évolution entre 2015 et 2021 est significative, il faut calculer la valeur suivante :

$$\frac{|\hat{Y}_{2021} - \hat{Y}_{2015}|}{\sqrt{se_{\hat{Y}_{2021}}^2 + se_{\hat{Y}_{2015}}^2}} \quad (4)$$

Entre 2015 et 2021, on obtient une valeur de 1.7 (inférieure à 1.96). Cela signifie que l'évolution du score moyen n'est pas statistiquement significative.

Les erreurs standards sont également calculées pour les répartitions dans les différents groupes de niveaux (tableaux 13 et 14).

TABLEAU 13 – Répartitions en % dans les groupes de niveaux - Compétences langagières et littératie Collège

Année	Groupe <1	Groupe 1	Groupe 2	Groupe 3	Groupe 4	Groupe 5
2015	3.1	11.9	27.0	30.0	18.0	10.0
2021	3.8	13.3	27.1	30.0	17.0	8.7

TABLEAU 14 – Erreurs standards des répartitions en % dans les groupes de niveaux - Compétences langagières et littératie Collège

Année	Groupe <1	Groupe 1	Groupe 2	Groupe 3	Groupe 4	Groupe 5
2015	0.3	0.6	0.7	0.7	0.6	0.5
2021	0.4	0.6	0.7	0.7	0.6	0.6

***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  $P$  :

$$D_{eff} = \frac{V_P(\hat{Y})}{V_{SAS}(\hat{Y})} \quad (5)$$

TABLEAU 15 – Effet du plan de sondage - Compétences langagières et littératie Collège

Année	Erreur Standard	Erreur SAS	<i>Design Effect</i>
2015	0.92	0.51	3.23
2021	1.25	0.7	3.19

Dans le cas d'un sondage en grappes, la précision est dégradée en comparaison d'un sondage aléatoire simple. Cela signifie qu'en 2021, un sondage aléatoire simple avec un effectif 3.19 fois moins important aurait conduit au même niveau de précision.

## 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) :

$$Y_i^j = \begin{cases} 1 & \text{si l'élève } i \text{ réussit l'item } j \\ 0 & \text{si l'élève } i \text{ échoue à l'item } j \end{cases} \quad (6)$$

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

$$p_j = \frac{1}{n} \sum_{i=1}^n Y_i^j \quad (7)$$

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  $i$  :

$$S_i = \sum_{j=1}^J Y_i^j \quad (8)$$

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 fidèle<sup>5</sup>.

Le coefficient  $\alpha$  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 :

$$\alpha = \frac{J\bar{r}}{1 + (J - 1)\bar{r}} \quad (9)$$

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 fidélité<sup>6</sup>.

### 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.

---

5. 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.

6. La littérature indique plutôt un seuil de 0,70 (Peterson, 1994). Cependant, comme le montre la formule ci-dessus, le coefficient  $\alpha$  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

$$r_{bis-point}(j) = corr(Y^j, S) = \frac{\bar{S}_{(j1)} - \bar{S}_{(j0)}}{\sigma_S} \sqrt{p_j(1 - p_j)} \quad (10)$$

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  $\sigma_S$  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)<sup>7</sup>. 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  $\phi$ , 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.

---

7. 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 :

$$y_{ij} = \begin{cases} 0 & \text{si } z_{ij} \leq \tau_j \\ 1 & \text{si } z_{ij} > \tau_j \end{cases} \quad (11)$$

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  $\tau_j$ , 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.

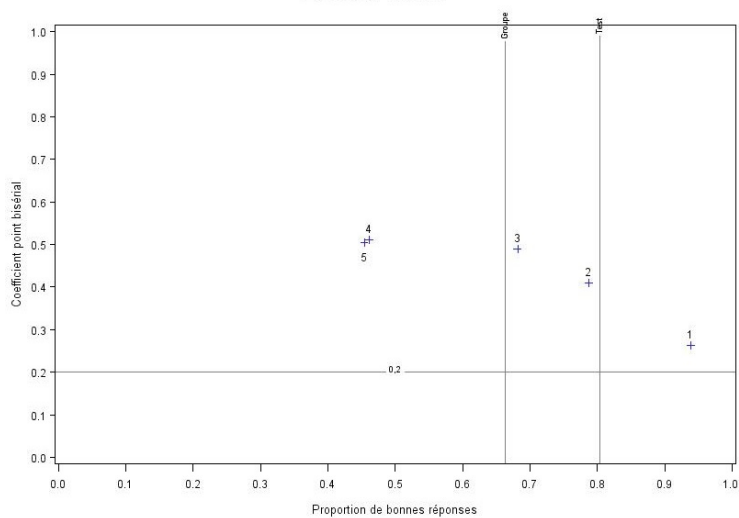
- Non-réponse terminale : l'élève s'est arrêté au cours de l'épreuve, potentiellement en raison d'un manque de temps. Des choix sont effectués pour déterminer le traitement de ces valeurs. Nous considérons que si les dernières réponses d'un élève sont vides alors il n'a pas eu le temps de les voir. Toutes les observations en non-réponses terminales sont enlevées des données. Les valeurs manquantes sont alors traitées de manière structurée. 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. 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 11). 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 11 – 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 bisérial 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 un prestataire extérieur (dans le cadre d'un marché public). Les réponses aux questions à choix multiples ont été numérisées et les codes de réponses stockés dans un fichier. Les questions ouvertes, demandant une rédaction plus ou moins longue de la part des élèves, ont été scannées 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'images (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 image, il arrive parfois qu'il y ait une différence de codage. Cette image 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 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-corrections. 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**

11 items, tous créés en 2021, ont été éliminés pour cause de rbis-point trop faible.

## 4 Modélisation

### 4.1 Méthodologie

#### 4.1.1 Modèle de réponse à l'item

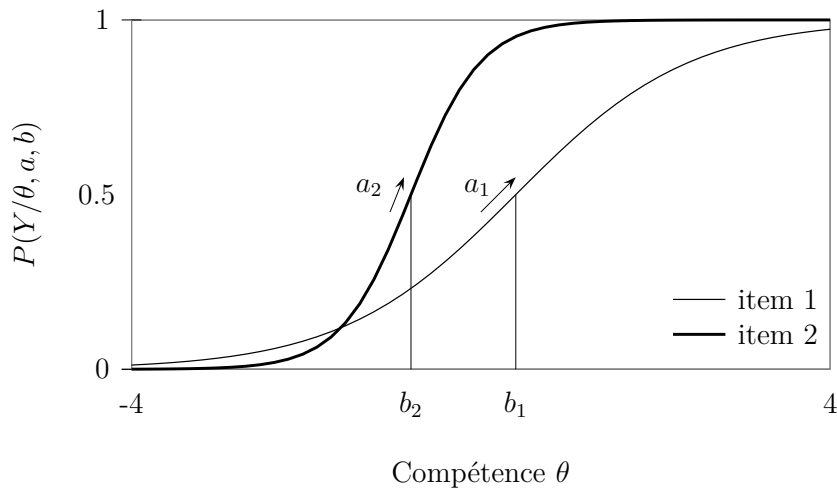
Le modèle de mesure utilisé est un modèle de réponse à l'item à deux paramètres avec une fonction de lien logistique (MRI 2PL) :

$$P_{ij} = P(Y_i^j = 1 | \theta_i, a_j, b_j) = \frac{e^{1,7a_j(\theta_i - b_j)}}{1 + e^{1,7a_j(\theta_i - b_j)}} \quad (12)$$

où la probabilité  $P_{ij}$  que l'élève  $i$  réussisse l'item  $j$  est fonction du niveau de compétence  $\theta_i$  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 12 représente les courbes caractéristiques de deux items selon cette modélisation.

FIGURE 12 – 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 12).

#### 4.1.2 Procédures d'estimation

L'estimation est conduite en deux temps : l'estimation des paramètres des items puis l'estimation des  $\theta$  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 (12) 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  $\theta$  pour les données du premier cycle (en l'occurrence, moyenne de 250 et écart-type de 50 pour l'année 2015).

Sous l'hypothèse d'indépendance locale des items<sup>8</sup>, la fonction de vraisemblance s'écrit :

$$L(\mathbf{y}, \xi, \theta) = \prod_{i=1}^n \prod_{j=1}^J P_{ij}^{y_{ij}} [1 - P_{ij}]^{1-y_{ij}} \quad (13)$$

---

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

où  $\mathbf{y}$  est le vecteur des réponses aux items (*pattern*),  $\xi$  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  $\theta$  est considérée comme une variable aléatoire de distribution connue, la probabilité inconditionnelle d'observer un *pattern*  $\mathbf{y}_i$  donné peut s'écrire :

$$P(\mathbf{y} = \mathbf{y}_i) = \int_{-\infty}^{+\infty} P(\mathbf{y} = \mathbf{y}_i | \theta_i) g(\theta_i) d\theta_i \quad (14)$$

avec  $g$  la densité de  $\theta$ .

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

$$L = \prod_{i=1}^n P(\mathbf{y} = \mathbf{y}_i) \quad (15)$$

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  $\theta$  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  $\theta_i$ , par exemple *via* la maximisation de la vraisemblance donnée par l'équation (13).

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 :

$$B(\theta_i^{(ML)}) = \frac{-J}{2I^2} \quad (16)$$

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 :

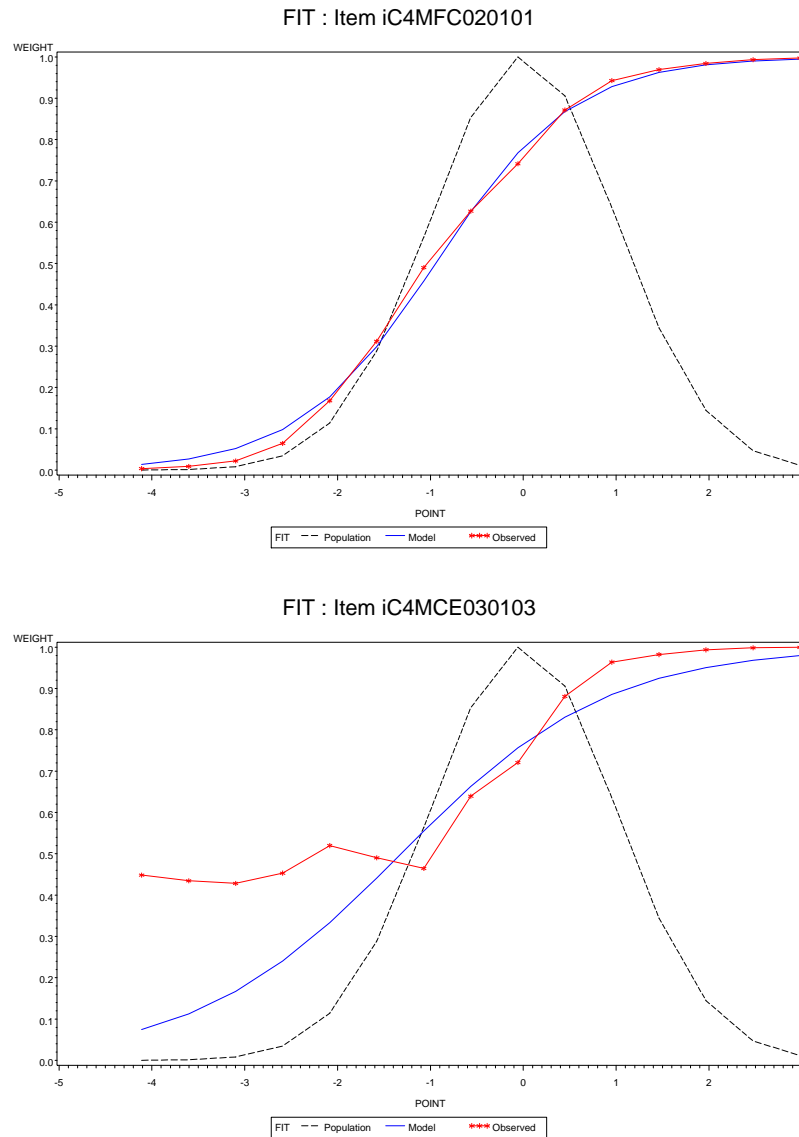
$$\frac{\partial \ln L}{\partial \theta_i} + \frac{J}{2I} = 0 \quad (17)$$

#### 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 13). 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 :

$$z_{ij} = \frac{Y_i^j - P_{ij}}{\sqrt{P_{ij}(1-P_{ij})}} \quad (18)$$

FIGURE 13 – 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  $\chi^2$ . L'indice *Infit* d'un item correspond à la moyenne pondérée des carrés des résidus, qui peut s'écrire :

$$Infit_j = \frac{1}{\sum_{i=1}^n w_{ij}} \sum_{i=1}^n w_{ij} z_{ij}^2 = \frac{1}{\sum_{i=1}^n w_{ij}} \sum_{i=1}^n (Y_i^j - P_{ij})^2 \quad (19)$$

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 :

$$P(Y | Z, G) \neq P(Y | Z) \quad (20)$$

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  $\beta$  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 :

$$I_j(\theta) = (1,7a_j)^2 P_j(\theta)(1 - P_j(\theta)) \quad (21)$$

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

L'information moyenne du test pour un élève de compétence  $\theta$  est la somme de l'information apporté par chaque item pour  $\theta$ . La courbe d'information du test est tracée pour un ensemble de valeurs de  $\theta$ . 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  $\theta$  des élèves est estimé.

## 4.2 Résultats

### 4.2.1 Identification des fonctionnements différentiels d'items (FDI)

17 items ont été éliminés des analyses pour cause de fonctionnements différentiels.

- 17 items d'ancrage 2015-2021

### 4.2.2 Bilan de l'analyse des items

Les évaluations CLL de 2015 et 2021 étaient constituées en tout de 249 items, répartis de la façon suivante :

- 88 items uniquement proposés en 2015
- 68 items d'ancrage papiers, proposés en 2015 et 2021
- 93 items nouveaux numériques uniquement proposés en 2021.

Cela représente 249 items passés par les élèves en tout, dont 162 en 2021.

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

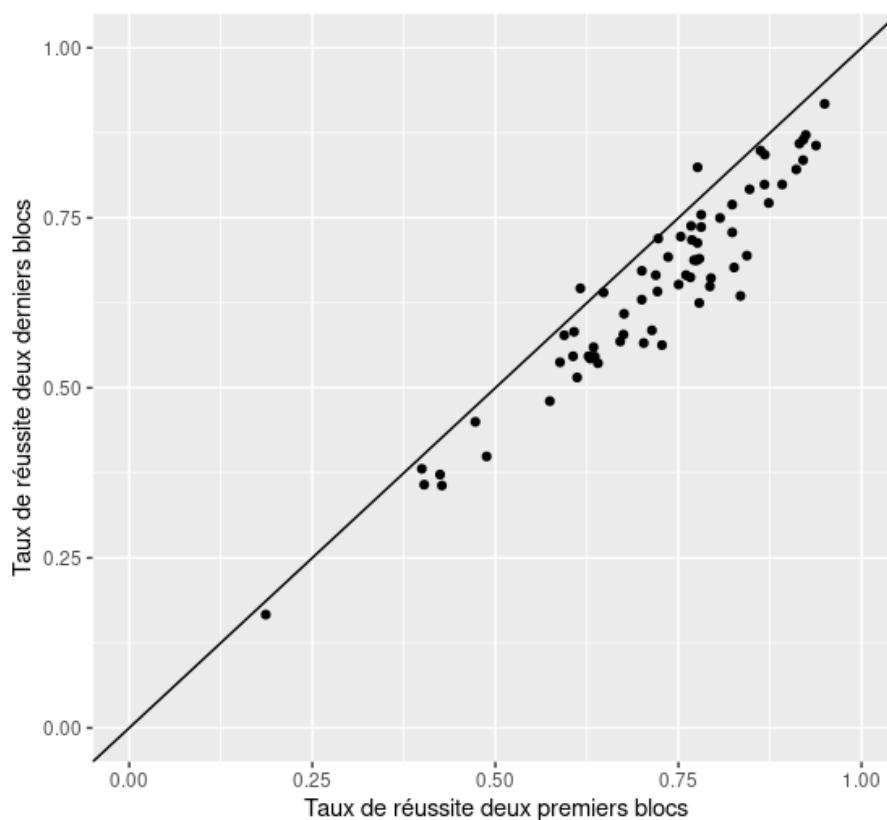
- 88 items uniquement présentés en 2015
- 45 items d'ancrages papiers, proposés en 2015 et 2021
- 88 items numériques nouveaux

221 items sont donc conservés dans l'analyse, dont 133 utilisés dans l'évaluation 2021.

### 4.2.3 Particularités de l'évaluation 2015

En 2015, l'évaluation se déroulait en une seule session de 1h45, tandis qu'en 2021, elle était divisée en deux séquences de 30 minutes sur deux jours distincts. Par ailleurs, une forte fatigue des élèves a été observée en 2015 (Figure 14), les performances étant nettement meilleures sur les items placés en début d'épreuve par rapport à ceux en fin de test. Afin d'harmoniser les conditions de passation entre 2015 et 2021, il a été décidé d'enlever les deux derniers blocs d'items de 2015 dans l'analyse. Comme le design de l'évaluation est tournant, tous les blocs sont concernés par cette suppression. Cela ne conduit pas à diminuer le nombre d'items pris en compte dans l'évaluation, mais à réduire le nombre de répondants pour chacun des items.

FIGURE 14 – Taux de réussite des items en 2015 en fonction de la position de leurs blocs



### 4.3 Calcul des scores

Comme indiqué précédemment, une analyse conjointe des données des 2 années a permis d'estimer les paramètres des items, puis les niveaux de compétences  $\theta$  des élèves. Afin de lever l'indétermination du modèle, la moyenne des  $\theta$  a été fixée à 250 et leur écart-type à 50, pour l'échantillon de 2015. Le tableau 16 présente les résultats obtenus.

TABLEAU 16 – Niveaux de compétences (moyennes des scores et écarts-types)  
Compétences langagières et littératie Collège

Année	Score moyen	Écart-type
2015	250	50
2021	247	49

## 5 Construction de l'échelle

### 5.1 Méthode

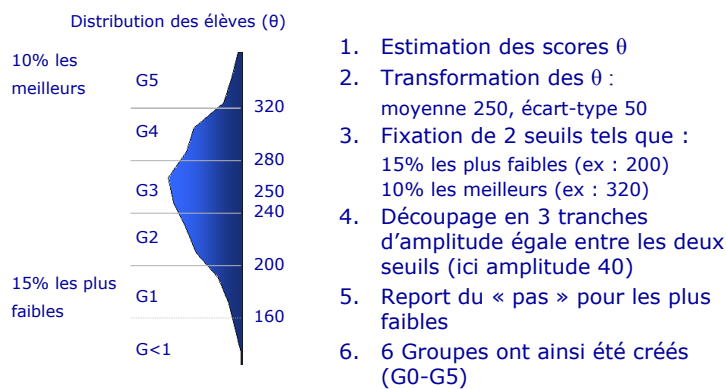
Les modèles de réponse à l'item permettent de positionner sur une même échelle les paramètres de difficulté des items et les niveaux de compétences des élèves. Cette correspondance permet de caractériser les compétences maîtrisées pour différents groupes d'élèves.

Les scores en Compétences langagières et littératie 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 2015. Puis, comme le montre la figure 15, 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 15 – Principes de construction de l'échelle



## 5.2 Caractérisation des groupes de niveaux

La mesure fine des acquis des élèves en Compétences Langagières et Littératie, objectif principal de ce CEDRE, s'observe grâce à l'échelle de performance qui sert de référence afin d'apprécier l'évolution de la répartition des élèves dans les différents groupes. Pour chaque groupe, elle précise le niveau global de maîtrise des connaissances et compétences. Les principaux résultats font l'objet d'une Note d'Information n° 22.29 (H. Durand de Monestrol, H. Rogie, 2022).

### **Groupe < 1 (3,8 % des élèves) Des difficultés potentielles dans des situations de lecture de vie courante**

Ces élèves sont potentiellement capables d'identifier un support fonctionnel de la vie courante à condition qu'il soit court et de nature familière. Ils peuvent retrouver une information dans des textes documentaires si celle-ci est rendue aisément repérable par sa situation dans le support ou des caractéristiques typographiques (données chiffrées, majuscules, etc.).

### **Groupe 1 (13,3 % des élèves) Des compétences élémentaires sur des supports très accessibles de la vie courante**

Ces élèves sont potentiellement capables de prélever une information dans un court texte si celle-ci est explicite et que le texte est repris de manière littérale dans la réponse attendue et/ou dans l'amorce. Confrontés à un contexte très familier, ils peuvent être à même d'exprimer une opinion, de dégager une idée principale, d'inférer une émotion ou le sens d'un mot.

**L'élève le plus faible de ce groupe a un score de 163.**

**Groupe 2 (27,1 % des élèves) Une maîtrise fragile des compétences de prélèvement, d'inférence ou de compréhension du sens global sur des supports scolairement familiers**

Ces élèves sont potentiellement capables de retrouver une information explicite reformulée dans un support contenant du vocabulaire spécialisé en rapport avec une discipline scolaire et d'en comprendre la visée. Ils peuvent être à même de rendre compte d'un raisonnement local ou d'établir un lien entre deux documents dans un contexte de vie courante. Confrontés à des supports non textuels ou discontinus, ils sont potentiellement capables de tirer une information d'un schéma ou d'un tableau à double-entrée. Face à un texte littéraire peu complexe, ils peuvent parvenir à inférer une émotion simple et dégager la connotation ou les idées principales.

**L'élève le plus faible de ce groupe a un score de 201.**

**Groupe 3 (30,0 % des élèves) Une capacité à réinvestir des connaissances pragmatiques au service de la compréhension et à comprendre une structure logique ou narrative**

Ces élèves sont potentiellement à même de réaliser des inférences dans un support contenant du vocabulaire spécialisé et de dégager des informations implicites en mettant en relation plusieurs documents. Ils sont potentiellement capables de s'appuyer sur leurs connaissances grammaticales pour accéder au sens d'un texte littéraire. Plus généralement, ils peuvent être à même de s'approprier la cohérence logique d'un support narratif, pour identifier un argument, repérer une relation cause-conséquence ou proposer une suite. Ils peuvent reconnaître une source fiable.

**L'élève le plus faible de ce groupe a un score de 240.**

**Groupe 4 (17,0 % des élèves) Une compréhension affinée de textes littéraires ou de nature scientifique donnant accès à des compétences d'esprit critique**

Ces élèves sont potentiellement capables de retrouver des informations explicites reformulées dans un texte littéraire complexe, d'identifier des nuances d'émotions chez les personnages ou de dégager des informations implicites en question ouverte. Ils sont potentiellement capables d'intégrer plusieurs informations issues d'un texte scientifique et de proposer des idées supplémentaires en cohérence avec ce dernier. Ils peuvent parvenir à évaluer les différences entre deux documents, à hiérarchiser des sources selon leur fiabilité.

**L'élève le plus faible de ce groupe a un score de 278.**

**Groupe 5 (8,7 % des élèves)****Un accès au sens de supports complexes et la mise en œuvre de stratégies permettant de hiérarchiser l'information**

Ces élèves sont potentiellement capables de retrouver des informations reformulées dans un support documentaire au vocabulaire très spécialisé, d'en tirer une conclusion ou d'interpréter les données d'un tableau. Ils peuvent parvenir à hiérarchiser des informations selon leur pertinence ou à rendre compte d'une vision fine et rigoureuse du sens global d'un support. Les élèves les plus performants sont potentiellement à même de hiérarchiser des documents selon leur fiabilité en s'appuyant spontanément sur la source de ces derniers.

**L'élève le plus faible de ce groupe a un score de 317.**

À chaque groupe d'élèves, sont associés des items de difficulté plus ou moins importante. Cela signifie que les élèves de ce groupe ont une probabilité d'au moins 50 % de répondre correctement à ces items. Suit un ensemble d'items représentatifs de chaque groupe.

**5.3 Exemples d'items****5.3.1 Item caractéristique du groupe < 1**

Le groupe < 1 est constitué des élèves les plus faibles : ils représentent 3,8 % des élèves. Ces élèves, susceptibles de rencontrer des difficultés y compris dans des situations de lecture de la vie courante, sont néanmoins capables de retrouver des informations aisément repérables.

Ainsi, dans un support assez dense comme l'article documentaire ci-dessous, publié dans National Geographic, ils peuvent être à même de prélever l'information demandée dans la première question.

FIGURE 16 – Support groupe &lt;1

**SITUATION 4 – LA GROTTTE CHAUVET**

Le pont d'Arc enjambe l'Ardèche depuis au moins 500 000 ans. Les hommes préhistoriques ont peut-être vu en lui le ventre bombé et la tête d'un mammouth ou d'un bison gardant l'entrée de la vallée, ce qui lui conféra une signification mythique. Environ vingt-cinq grottes ornées ont été découvertes ici depuis la fin du XIX<sup>e</sup> siècle. Pourtant, jusqu'à la mise au jour de la grotte Chauvet, en 1994, l'art pariétal de l'Ardèche est resté peu connu si on le compare à celui de Lascaux (20 000 ans) ou d'Altamira (17 000 ans).

Les premières photos captivèrent tant les experts que le public. Jusque là, les spécialistes s'en étaient tenus à la théorie selon laquelle l'art avait progressé par étapes, lentement, des dessins primitifs aux figurations naturalistes plus tardives. L'utilisation de la perspective, l'emploi de l'estompe pour rendre les ombres et l'élégance des dessins de Chauvet ne pouvaient que constituer l'aboutissement de ce processus. Cependant, après les premières datations au carbone 14, les préhistoriens durent admettre que ces fresques étaient près de deux fois plus anciennes que celles des grottes les plus célèbres. Elles ne correspondaient pas à l'apogée de l'art préhistorique mais à ses tout premiers débuts connus.

On dispose à ce jour de trente datations au radiocarbone de Chauvet, plus que pour aucune autre grotte ornée. Elles montrent que cette grotte était utilisée il y a environ 35 000 ans et que dix-sept compositions ont été exécutées à cette époque. Une deuxième vague de visiteurs est attestée, datant d'environ 30 000 ans, mais on ignore encore si elle a laissé des peintures. Devant le caractère exceptionnel de cette découverte, le ministère de la Culture a confié l'étude à une équipe entièrement professionnelle, une première pour un site pariétal. Deux ans et près de 15 millions de francs ont été nécessaires pour procéder à divers aménagements, comme élargir l'entrée, installer un système de sécurité et aménager un système de passerelles, sans altérer le délicat équilibre naturel de la grotte.

Les douze membres de mon équipe étudient les peintures des parois et les traces humaines et animales au sol, avec le concours, entre autres, de spécialistes de la datation, de l'analyse des pigments et du comportement animal. Nous bénéficions aussi des conseils de grands spécialistes de l'art pariétal de dix pays différents. Vingt-huit auteurs - géologues, archéologues, anthropologues et historiens de l'art - se sont joints à moi pour écrire le premier ouvrage consacré au site\*.

Les hommes n'ont jamais habité la grotte. Des artistes, des enfants, peut-être des adeptes de certains rites, sont entrés dans ce sanctuaire souterrain. Puis une partie de la falaise surplombant l'entrée s'est effondrée, obturant la grotte et préservant ses remarquables œuvres d'art.

Article paru dans la revue © « National Géographic », août 2001

\* *La grotte Chauvet. L'Art des origines*, sous la direction de Jean Clottes, a été publié en mai 2001 par les éditions de Seuil © Tous droits réservés

*Quand la grotte Chauvet a-t-elle été mise au jour ?*

<input checked="" type="checkbox"/> en 1994	93,9%
<input type="checkbox"/> il y a 17 000 ans	1,2%
<input type="checkbox"/> il y a 20 000 ans	1,1%
<input type="checkbox"/> en 2001	3%

**Analyse de la question**

Le repérage de l'information, qui se trouve dès le premier paragraphe, est facilité par la majuscule au nom « Chauvet », par l'écriture chiffrée de l'information attendue et enfin par la reprise littérale du texte dans la question avec l'expression « mise au jour ». Ainsi, si l'information est repérée, rien ne garantit qu'elle soit comprise.

**Analyse des erreurs**

On peut noter que la proposition de réponse erronée la plus attractive a été la date de publication de l'article, information saillante présente deux fois en paratexte, qui aura, à la faveur d'une confusion entre « publier » et « mettre au jour », probablement attiré des élèves n'ayant pas repéré la fin du premier paragraphe. En effet, les informations chiffrées situées quant à elles en fin de premier paragraphe et qui constituent les deux autres distracteurs, ont été peu choisies.

**Taux de réussite et non réponse**

Réussite	Non-réponse
93,9 %	1,1 %

**5.3.2 Item caractéristique du groupe 1**

Le groupe 1 représente 13,3 % des élèves. Il est constitué des élèves témoignant de compétences élémentaires dans des situations de vie courante : ils montrent une aptitude à procéder à des inférences élémentaires et à accéder au sens global dès lors que le support est accessible par sa brièveté, son niveau de langue et sa thématique. Écrans et vie quotidienne est un groupement de documents très accessibles autour de l'utilisation des écrans dans la vie de tous les jours. Ce groupement thématique a été proposé dans la séquence sur ordinateur : les trois documents qui le composent sont répartis dans trois onglets d'une « liseuse ». La quatrième question porte sur le troisième document : un très court article de journal intitulé : « Comment échapper à l'overdose ». Les élèves sont interrogés sur l'image qui est donnée des écrans à travers cet article.

FIGURE 17 – Support groupe 1

ÉCRANS ET VIE QUOTIDIENNE

1
2
3

**Document 3**

COMMENT ÉCHAPPER À L'OVERDOSE		
<p><b>Évitez l'addiction.</b> Il faut s'interroger sur les raisons qui poussent à allumer les écrans. Ne pas trop solliciter les autres par écrans interposés : se rencontrer plutôt que se téléphoner, se téléphoner plutôt qu'écrire des e-mails...</p> <p><b>Luttez contre le stress.</b> Ne vous laissez pas interrompre</p>	<p>tout le temps dans vos tâches importantes. Ménagez-vous des moments d'indisponibilité ; enclenchez une notification automatique d'absence.</p> <p><b>Faites une pause.</b> Face à l'avalanche numérique, il faut réapprendre à ne rien faire. Pour cesser d'être asphyxié, Annulez vos alertes sonores</p>	<p>pour ne pas être incité à regarder votre écran en permanence.</p> <p><b>Montrez l'exemple aux enfants.</b> Passez un contrat avec eux pour définir une durée autorisée devant l'écran. Chaque soir, lisez-leur une histoire ou le chapitre d'un livre pour éviter le « zapping ».</p>

Précédent
Page 3 / 3
Suivant

**Question 4/8**

**Le document 3 présente les écrans comme...**

une obligation moderne.

un outil révolutionnaire.

un remède anti-stress.

une véritable drogue.

- |  |       |
|--|-------|
| <input type="checkbox"/> une obligation moderne          | 3,5%  |
| <input type="checkbox"/> un outil révolutionnaire        | 2,8%  |
| <input type="checkbox"/> un remède anti-stress           | 7,6%  |
| <input checked="" type="checkbox"/> une véritable drogue | 85,7% |

#### Analyse de la question

La réponse attendue, « une véritable drogue », est inférable à partir du titre principal mais également du premier intertitre : « évitez l'addiction ». On peut noter que la dénonciation de la surutilisation des écrans est un discours très relayé et donc certainement familier pour la plupart des élèves : c'est un facteur facilitant.

### Analyse des erreurs

L'ensemble des distracteurs proposent cependant des idées plausibles qui reposent sur les éventuelles représentations personnelles des élèves. La proposition de réponse erronée la plus efficace est « un remède anti-stress » qui témoigne probablement d'une lecture morcelée s'appuyant sur le 2e intertitre : « lutez contre le stress », encadré en rouge sur la diapo. Par ailleurs, les distracteurs « un outil révolutionnaire » et « un remède anti-stress » s'opposent à la réponse correcte par leur connotation positive : ils témoignent donc tous d'un contresens sur la visée générale du support. La réponse « une obligation moderne » est contradictoire avec le discours de l'article, encourageant à se détacher des écrans.

### Taux de réussite et non réponse

Réussite	Non-réponse
85,7 %	0,4 %

#### 5.3.3 Item caractéristique du groupe 2

Le groupe 2 représente 27,1 % des élèves. Ces élèves sont potentiellement capables d'accéder aux sens de textes littéraires ou documentaires et à réaliser les tâches inhérentes à l'étude de document : prélever des informations explicites, mêmes non littérales, réaliser des inférences et rendre compte du sens global en s'appuyant sur le texte. Voici un texte narratif de longueur moyenne, d'un niveau de langue courant et écrit au présent de narration. Il s'agit d'un extrait du récit autobiographique de Joseph Joffo intitulé *Un sac de billes*, publié en 1973. La question 3 porte sur la compréhension de la chaîne référentielle.

## FIGURE 18 – Support groupe 2

## SITUATION 4 – UN SAC DE BILLES

- C'est le crissement des freins qui m'a fait retourner, malgré la porte fermée, le son a résonné, strident.  
À travers les rideaux, j'ai vu les deux camions bloquer les rues en travers.  
— Regardez !...
- 5 Il n'y a pas besoin que je leur dise, tous les hommes se sont tus et regardent descendre des soldats noirs aux bérets inclinés. Ceux-là sont les plus détestés, ce sont les chasseurs de résistants, les miliciens.  
Mitraillette au ventre, j'en vois un courir vers la ruelle Saint-Jean ; ils connaissent le pays, ils doivent savoir que l'on peut s'enfuir par là, en passant par les prés.
- 10 Il y en a un au centre qui fait de grands signes et j'en vois quatre se diriger vers nous.  
— Ils viennent ici, dit le patron.  
— Petit...
- 15 Je me retourne et regarde l'homme qui a parlé : c'est un des buveurs, mais je ne l'ai jamais vu, un petit homme assez vieux, habillé en velours sombre, il me sourit tranquille et vient vers moi.  
On ne nous voit pas de la rue.  
Une enveloppe froissée tombe dans ma sacoche. Il l'a sortie de sa poche. Il jette dessus le journal qu'il tenait à la main.  
Les miliciens poussent la porte.
- 20 Les lèvres du petit homme ne bougent pas, pourtant il parle. Ses yeux ne me regardent plus, mais c'est à moi qu'il s'adresse.  
— M. Jean, dit-il, au Cheval-Blanc.  
Sa main me pousse vers la porte. Je sors et me heurte à deux torsos sombres sanglés de baudriers. Ce sont deux types bronzés, l'ombre de leurs bérets cache leurs yeux.
- 25 — Les pattes en l'air, vite.  
Le plus maigre bondit comme un chat, c'est un super énervé, de la hanche, il balance une carafe qui tombe à terre et le canon de sa mitrailleuse trace des zigzags. Le patron ouvre la bouche, je vois une main hâlée l'empoigner par le devant de la chemise et le pousser contre le comptoir.
- 30 Le deuxième milicien me regarde et a un geste du pouce derrière son épaule.  
— Tire-toi, gamin.  
Je passe entre les deux hommes, ma sacoche sous le bras. De la vraie dynamite, cette sacoche, mais la dynamite n'explose pas toujours. Qui ferait attention à un petit livreur de journaux ? Je réfléchis en pédalant. Qui est ce petit homme en velours ?
- 35 Au coin de la place, je me retourne.  
Il est là, entre deux miliciens, il a ses mains sur la tête. Il est loin, peut-être est-ce une grimace due à un rayon trop vif du soleil, mais j'ai l'impression qu'il sourit et que ce sourire est pour moi. Au Cheval-Blanc à présent, vite.

Joseph JOFFO, *Un sac de billes*,  
© Ed Gallimard Folio, tous droits réservés

*Le plus maigre (ligne 26) remplace...*

- |  |       |
|--|-------|
| <input checked="" type="checkbox"/> un milicien    | 77,5% |
| <input type="checkbox"/> le narrateur              | 2,6%  |
| <input type="checkbox"/> Monsieur Jean             | 9,5%  |
| <input type="checkbox"/> le petit homme en velours | 8,9%  |

**Analyse de la question**

Pour faciliter le repérage, l'expression « le plus maigre » a été soulignée dans le texte (ligne 26). L'élève qui a le plus de chance de répondre à cette question est celui qui aura réussi à se faire une représentation mentale du milicien en retenant l'association « miliciens / mitrailleuse » donnée en début de texte (lignes 7-8). Il peut alors mettre en relation cette image avec la mention de la mitrailleuse

dans la phrase commençant par « le plus maigre ».

Au niveau de la chaîne référentielle, la mention littérale des miliciens est plutôt lointaine : 6 lignes en amont de l'expression soulignée. Les deux reprises nominales qui séparent le groupe nominal « les miliciens » (ligne 19) de « le plus maigre » (ligne 26) sont peu transparentes prises isolément : « deux torsos sombres sanglés de baudriers » (ligne 23) puis « deux types bronzés » (ligne 24). L'inférence peut aussi éventuellement se faire avec l'appui en aval sur la référence au « deuxième milicien » (ligne 30).

### Analyse des erreurs

Le choix d'un des distracteurs révèle une méconnaissance ou une mise en œuvre inefficace des stratégies de lecture propres au texte narratif. La réponse « le narrateur », qui a attiré 2,6 % des élèves, est le signe d'une incompréhension du système énonciatif d'un texte narratif à la première personne. « Monsieur Jean », le distracteur le plus choisi, relève sans doute d'une lecture morcelée sans véritable compréhension des événements : c'est la référence la plus saillante et la plus proche de l'expression « le plus maigre ». Le choix de la proposition « le petit homme en velours », personnage principal de l'extrait en dehors du narrateur, montre que les événements narrés n'ont pas été compris.

### Taux de réussite et non réponse

Réussite	Non-réponse
77,5 %	1,5 %

#### 5.3.4 Item caractéristique du groupe 3






Le groupe 3 représente 30,0 % des élèves. Ces élèves parviennent à effectuer des tâches locales sur des supports moins accessibles que le groupe précédent et à mettre en relation les informations de supports utilisant un niveau de langue courant. Ils sont également capables de faire appel à leurs connaissances grammaticales pour réaliser des inférences et parviennent à se prononcer sur la fiabilité d'une source. Le support ci-dessous est un exemple des documents qualifiés de « numériques » intégrés à l'évaluation en 2021. Il s'agit d'un site fictif proposant des hébergements dits « insolites ». Sur le modèle des évaluations internationales comme PISA, le document est introduit par une courte mise en situation : « Jérémy a entendu parler d'hébergement insolite mais il ne sait pas de quoi il s'agit. Il cherche donc sur Internet et une page d'accueil s'affiche ». Les premières questions invitent les élèves à prélever ou à croiser des informations concernant les différents hébergements. La question 4, en format ouvert, demande à l'élève de proposer une définition de « hébergement insolite ». Comme dans les évaluations internationales de compréhension de l'écrit, l'orthographe et la syntaxe ne

sont pas prises en compte et l'élève n'est pas dans l'obligation de produire une réponse rédigée. On peut noter par ailleurs noter le taux de non réponse élevé, sachant que des réponses du type « je ne sais pas » sont comptabilisés comme des réponses incorrectes et n'alimentent pas ce taux de non réponse.

FIGURE 19 – Support et question groupe 3

HÉBERGEMENTS

Jérémy a entendu parler d'hébergement insolite mais il ne sait pas de quoi il s'agit. Il cherche donc sur internet et une page d'accueil s'affiche.

Hébergements insolites en France					
destinations	dates - / - / -	- personnes	thèmes	<a href="#" style="color: red; text-decoration: none;">nos coups de cœur du mois</a>	contact
<b>Le zoo</b>			Tout le charme de l'Afrique en France, sans crainte des prédateurs car de grandes baies vitrées protègent les hôtes. 24 cabanes sur pilotis avec tout le confort d'un hôtel. Compter une journée complète pour tester toutes les animations. À partir de 140 € par personne avec petit déjeuner, dîner et entrée du parc animalier. <a href="http://www.hutte_au_zoo.com">http://www.hutte_au_zoo.com</a>		
<b>Le train</b>			Aux portes d'une grande ville, voici des wagons et une locomotive qui n'attendent que vous . Couchage pour 8 personnes dans l'un ou pour 4 personnes dans la locomotive. Un simulateur vous passera sur la vitre le défilement du trajet selon votre choix : Orient - Express , Indian Pacific, Transsibérien ou Victoria Express. À partir de 130 € pour 2 et par nuit selon la période de l'année. <a href="http://www.Amourdestrains.com">http://www.Amourdestrains.com</a>		
<b>Le château d'eau</b>			Chambres aveugles mais inconfort largement compensé par une vue magnifique au sommet sur les montagnes environnantes ; terrasse aménagée avec transats, jacuzzi et sauna ; garde-fou de sécurité. Grand confort et décoration soignée des chambres. 140 € pour deux en semaine, petits déjeuners inclus. Week-end promotionnels en novembre : 170 € pour un couple avec deux enfants. <a href="http://www.vuesuralpes.com">http://www.vuesuralpes.com</a>		
<b>La roulotte</b>			40 roulottes de toutes les tailles, vous tendent leurs bras pour héberger de 2 à 5 personnes. Elles s'éparpillent dans un parc de 100 hectares au milieu des ânes et des chevaux. De nombreuses activités en lien avec les équidés vous sont également proposées (payantes). De 80 à 150 € la nuit selon le nombre de personnes <a href="http://www.ane-roulotte.fr">http://www.ane-roulotte.fr</a>		
<b>Le tonneau</b>			On peut y cuisiner, s'y laver, y dormir dans des lits très confortables. Par contre, si vous faites plus d'1,90 m, gare à votre tête quand vous rasez les murs ! Situé dans une ferme, goûtez les produits locaux avec les propriétaires en choisissant la pension complète. Vous pouvez aussi vous contenter du petit déjeuner si vous louez à la nuitée. 90€ pour 4 personnes, 120 € pour 6, petit déjeuner compris. <a href="http://www.diogeneetmoi.com">http://www.diogeneetmoi.com</a>		

**Question 4/4**

Après avoir consulté le site, quelle définition donneriez-vous de "hébergement insolite" ?

### Analyse de la question

Pour répondre à cette question, les élèves devaient identifier le point commun entre les hébergements proposés sur le site fictif et donc mettre en relation des informations pour comprendre le sens global. Ainsi, pour être validée, la réponse devait expliciter l'adjectif « insolite » par l'idée d'étrangeté ou de caractère inhabituel. L'élève pouvait proposer un synonyme (premier exemple de réponse ci-dessous) ou une paraphrase (deuxième exemple de réponse). Une réponse plus rédigée (troisième exemple) n'est pas valorisée : elle est comptabilisée comme une réponse correcte au même titre que les autres. De plus l'orthographe et la grammaire ne sont pas évaluées ici.

### Exemples de réponses d'élèves codée en réponse correcte

*bizarre*

*être hébergé dans des endroits ou on n'y aurait jamais pensé*

*C'est un hébergement insolite hors du commun, original, atypique, qui sortirait dans tous les cas des clichés auxquels les voyageurs sont habitués à trouver dans différents pays du monde.*

### Analyse des erreurs

Les réponses incorrectes sont de plusieurs ordres. Certains élèves ne mettent pas du tout les informations en relation et désignent un des hébergements proposés dans le site. La réponse de ce type la plus courante cible le zoo, le premier hébergement de la liste, mais les autres hébergements sont également représentés (premier exemple de réponse ci-dessous). D'autres élèves échouent dans la mise en relation et prélèvent des caractéristiques saillantes de certains hébergements (deuxième exemple). On peut noter que la belle vue offerte par le château d'eau était mise en avant par une des questions précédentes. La réponse donnée en troisième exemple ci-dessous rentre en contradiction avec certaines informations du support et relève sans doute des représentations personnelles de l'élève. Ce type de réponse a pu être influencé par une question précédente interrogeant sur le logement le plus avantageux financièrement. On trouve également des réponses hors-sujet comme dans le quatrième exemple qui explicite le mot « hébergement » sans qu'on puisse vérifier la compréhension du mot « insolite ». Certains élèves disposant d'un bagage lexical pauvre peuvent avoir du mal à proposer un équivalent assez précis : il est impossible de savoir si « nouveau » dans le cinquième exemple, est dans l'esprit de l'élève, l'équivalent d'original ou inédit (auquel cas l'élève aurait réussi la tâche attendue) ou simplement celui de récent ou neuf. Dans le doute, la réponse est considérée comme trop vague. Enfin, le dernier exemple de réponse incorrecte est une simple paraphrase de la question posée.

**Exemples de réponses d'élèves codées en réponse incorrecte**

*Le zoo*

*Il est pratique et peut donner une belle vue sur le nature*

*c'est pas chère*

*C'est comme un hotel pour passer les vacances*

*un hebergement nouveau*

*c'est un lieu d'hebergement qui est aménager dans des lieu insolites*

**Taux de réussite et non réponse**

Réussite	Non-réponse
62,8 %	16 %

**5.3.5 Item caractéristique du groupe 4**

Le groupe 4 rassemble 17 % des élèves. Ces élèves sont capables de mettre en œuvre les différentes compétences de compréhension de l'écrit face à des supports littéraires de langue soutenue ou des documents comportant du vocabulaire spécialisé. Ils montrent une compréhension fine des émotions ou des intentions des personnages dans des textes littéraires et détectent des divergences entre différents documents. Ils peuvent rendre compte de l'organisation d'un propos en passant du concret à l'abstrait. Le support ci-dessous est encore un exemple des documents qualifiés de « numériques » intégrés à l'évaluation en 2021 : il s'agit d'un échange fictif de courriels entre une femme à la recherche d'un appartement et un agent immobilier. Dans la question 2, l'élève est invité à rendre compte de l'organisation du propos du premier courriel en remettant dans l'ordre dans la deuxième colonne les quatre étiquettes fournies.

FIGURE 20 – Support et question groupe 4

APPARTEMENT

Loubna Khacer vit seule à Libourne avec sa fille de cinq ans. Elle vient de décrocher un emploi au centre commercial de Mériadeck à Bordeaux. Ne possédant pas de voiture, Loubna souhaite se rapprocher de son lieu de travail. Elle contacte Thierry Demestre qui lui propose plusieurs appartements à la location.

1
2

De : Loubna Khacer < LoubnaKhacer@tospeed.fr > Envoyé : Mer 25/07/2017 9:05 À : Thierry Demestre < T.Demestre@tospeed.fr > Objet : Demande de location Monsieur, J'ai appris par Christophe Lebas, votre neveu, que vous proposiez plusieurs appartements à la location à Bordeaux. Je suis à la recherche d'un logement de préférence situé dans le secteur du centre commercial. Je souhaiterais retrouver un appartement disposant de deux chambres, situé de préférence au rez-de-chaussée ou accessible par ascenseur, pour un loyer maximal de 800 euros charges comprises. J'aimerais savoir si vous aviez encore des biens de ce type disponibles et si vous pouviez, dans ce cas, me transmettre un descriptif incluant le montant du loyer. Je dois débiter mon nouvel emploi le 15 septembre prochain et aimerais donc déménager au plus tard le 25 Août. Cordialement.

Précédent
Page 1 / 2
Suivant

**Question 2/4**

**Remettre dans l'ordre les différentes étapes de la demande de Loubna (page 1).**

- détail des attentes
- secteur souhaité
- délais à respecter
- demande d'informations

>

### Analyse de la question

Chaque expression proposée dans les étiquettes à classer correspond à un paragraphe du courriel et présente de manière abstraite le contenu de ce paragraphe. Dans le cas du premier paragraphe, l'association est facilitée par la reprise du mot « secteur ». Pour les autres paragraphes, il n'y a pas de correspondance littérale entre le contenu du paragraphe et l'expression correspondante dans les étiquettes. On peut cependant noter que les amorces des paragraphes 2 et 3

« je souhaiterais retrouver » d'une part, et « j'aimerais savoir » d'autre part, renvoient de manière assez transparente respectivement aux attentes et à la demande d'informations. Quant au paragraphe sur les délais, les dates sont des informations saillantes qui peuvent permettre de l'identifier.

#### **Analyse des erreurs**

Il est probable que les élèves ayant échoué à cette question aient davantage été déstabilisés par le maniement des expressions abstraites que par la compréhension du courriel en lui-même.

#### **Taux de réussite et non réponse**

Réussite	Non-réponse
45,6%	1,0 %

#### **5.3.6 Item caractéristique du groupe 5**

Le groupe 5 rassemble 8,7 % des élèves. Ces élèves parviennent à s'appropriier les documents les plus complexes parmi ceux qui leur sont proposés et sont capables de hiérarchiser les informations selon leur pertinence. Ils savent prendre d'eux-mêmes appui sur la source d'un document pour évaluer sa fiabilité.

Le groupement de documents ci-dessous porte sur les effets des écrans sur la santé. La question 7 est une question ouverte par laquelle les élèves sont appelés à se prononcer sur la fiabilité relative des documents 1 et 3 et à justifier leur réponse.

FIGURE 21 – Support et question groupe 5

**Document 1**

D'après un psychiatre travaillant sur les troubles du sommeil, la luminosité, captée par la rétine, éveille et stimule le cerveau. La lumière bloque la production de mélatonine, « hormone du sommeil », qui donne l'heure biologique à notre cerveau et à notre corps. Une exposition à la lumière des écrans retarde l'endormissement et décale nos rythmes naturels.

Des recherches récentes révèlent une autre répercussion de l'abus des écrans sur notre santé : les signes de puberté, par exemple les premières règles, apparaissent plus tôt. En effet, la période de puberté est aussi régulée en grande partie par l'hormone du sommeil, la mélatonine, qui augmente lorsque la luminosité naturelle diminue, mais les enfants qui regardent beaucoup les écrans produisent moins de mélatonine à cause de la lumière artificielle que ces écrans projettent.

L'Académie des sciences\* liste âge par âge les risques psychologiques encourus par une utilisation abusive des écrans : retard de langage, déficit d'attention, appauvrissement de la mémoire, de la capacité de synthèse et d'apprentissage.

Page 1 / 3

**Document 2**

La glande pinéale commence à sécréter de la mélatonine

Pic de mélatonine au milieu de la nuit

Baisse de la mélatonine durant la journée

Source : <http://nouveaux.technologies-et-sommeil.com/le-monteil.com>

Page 2 / 3

**Document 3**

**COMMENT L'EXPOSITION À LA LUMIÈRE BLEUE AFFECTE VOTRE CERVEAU ET LE CORPS**

EN PERTURBANT LA MÉLATONINE, LA LUMIÈRE DU SMARTPHONE POUVEZ LES HORAIRES DE SOMMEIL. CELA CONDUIT À TOUTES SORTES DE PROBLÈMES DE SANTÉ.

Le perturbateur de vos habitudes de sommeil peut vous distraire et retarder votre réveil le lendemain.

De plus, l'usage excessif de votre téléphone peut vous empêcher de dormir et perturber votre sommeil.

Il y a des preuves scientifiques montrant que l'exposition à la lumière bleue pendant la nuit peut perturber votre sommeil et affecter votre santé.

Les chercheurs ont découvert que l'exposition à la lumière bleue pendant la nuit peut perturber votre sommeil et affecter votre santé.

Source : <https://sain-et-naturel.com/lumiere-bleue-du-smartphone-affecte-cerveau-corps.html>

Page 3 / 3

**Question 7/7**

Quel est le document qui inspire le plus confiance entre le document 1 et le document 3 ? Expliquez la raison de votre choix.

### Analyse de la question

Pour que la réponse soit validée, l'élève devait identifier le document 1 comme étant le plus fiable et s'appuyer sur des éléments pertinents pour justifier ce choix, à savoir la source du document ou les références scientifiques qu'il contient. Le premier exemple de réponse correcte ci-dessous s'appuie sur une des sources citées dans l'article à savoir une psychiatre. Les élèves pouvaient également s'appuyer sur la référence à l'Académie des sciences pour justifier la fiabilité des informations. Le deuxième exemple s'appuie sur la source du document lui-même. Une comparaison explicite avec le document 3 n'était pas exigée même si elle est présente dans les réponses les plus développées, à l'image du troisième exemple.

### Exemples de réponses d'élèves codées en réponse correcte

*C'est le document 1 car ce sont expliqués par un psychiatre donc c'est un professionnel.*

*Pour moi c'est le document 1 car c'est un extrait de journal donc l'auteur à vérifier ses informations avant de les publiés.*

*Selon moi, il s'agit du document 1 car celui-ci s'appuie sur le témoignage d'un psychiatre travaillant sur les troubles du sommeil, il est aussi cité les recherches de l'Académie des sciences, tandis que le document 3 ne cite aucune source venant d'un professionnel mis à part l'URL du site en question, on a donc aucune preuve que ce document soit véridique puisque aucune trace des éléments cités ne sont inscrit.*

### **Analyse des erreurs**

Les types de réponses incorrectes données le plus fréquemment à cette question sont la réponse sans justification, quel que soit le document choisi (premier exemple de réponse incorrecte) ou la justification inappropriée. Les élèves ont alors tendance à s'appuyer sur la forme et l'accessibilité du document : ils plébiscitent les images et les couleurs pour le document 3 (cf. deuxième exemple ci-dessous) ou déclarent au contraire préférer le texte continu en cas de choix du document 1 (cf. troisième exemple ci-dessous). Évoquer les propos du document constitue également une justification inappropriée. Les élèves semblent évaluer leur pertinence, confondant ainsi fiabilité et utilité : c'est le cas du quatrième exemple de réponse incorrecte. Enfin, certains élèves, tout en ayant conscience qu'il existe un lien entre la confiance que l'on peut accorder à un document et sa source, ne parviennent pas à s'en servir de manière pertinente comme en témoigne le dernier exemple de réponse incorrecte.

### **Exemples de réponses d'élèves codées en réponse incorrecte**

*1*

*3 car avec des images on comprend mieux*

*c est le 1 parce que je trouve que en version écrite cela fait moïn peur*

*c est le 1 car il parle se comment reussir a dormir*

*3 car il explique et indique leur source*

### **Taux de réussite et non réponse**

Réussite	Non-réponse
19,5 %	11,6 %

## 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 (H. de Monestrol & H. Rogie, 2021).

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 établissements, et non au niveau individuel. En effet, en 2021, la PCS des parents est disponible pour chaque élève, mais elle ne l'était pas dans le cycle antérieur. Pour chaque établissement des échantillons de 2015 et 2021, la moyenne de l'indice de position sociale a été calculée et la population a ensuite été découpée en quatre groupes selon les quartiles (tableau 6.1).

Indice moyen de l'étab.	Année	Score moyen	Ecart type
<b>Groupe 1 (25% les plus défavorisés)</b>	2015	234	49
	2021	230	50
<b>Groupe 2</b>	2015	249	48
	2021	248	49
<b>Groupe 3</b>	2015	252	49
	2021	<b>248</b>	46
<b>Groupe 4 (25% les plus favorisés)</b>	2015	264	50
	2021	260	48

Note de lecture : en 2021, le score moyen des élèves appartenant au quart des collègues les plus favorisés (Groupe 4) diminue de 4 points par rapport à 2015. Les évolutions significatives sont indiquées en gras.

## 6.2 Élaboration des questionnaires de contexte

Pour pouvoir enrichir l'analyse des résultats, un questionnaire de contexte a été élaboré. Un questionnaire élève a été ajouté à la fin du cahier d'évaluation. Composé de 26 affirmations avec lesquelles les élèves devaient se déclarer « pas du tout d'accord », « pas vraiment d'accord », « un peu d'accord » ou « tout à fait d'accord », le questionnaire avait principalement pour but d'interroger leurs stratégies de lecture. La consigne demandait aux élèves d'exprimer leur opinion sur l'étude de documents et les plaçait donc dans un contexte scolaire. Les affirmations étaient présentées dans un ordre aléatoire.

## 6.3 Motivation des élèves face à la situation d'évaluation

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 22) a été introduit dans plusieurs évaluations conduites au niveau national par la DEPP, y compris dans CEDRE Compétences langagières et littéraires. 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 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.

L'élève est interrogé sur sa motivation au début de chacune des séquences (papier et numérique). Puis, à la fin de chaque épreuve, il doit répondre à une question sur la difficulté perçue ainsi qu'à deux questions sur son degré d'application : pendant le test qu'il vient d'effectuer d'une part et dans le cas d'évaluations prises en compte dans le bulletin scolaire d'autre part.

TABLEAU 17 – Quel est votre degré de motivation pour faire cette évaluation ?  
(CEDRE Compétences langagières et littératie Collège)

<b>Avant le test :</b>	<b>%</b>
<i>Je suis très motivé(e)</i>	8,4
<i>Je suis motivé(e)</i>	41,3
<i>Je suis peu motivé(e)</i>	34,9
<i>Je ne suis pas du tout motivé(e)</i>	15,4

TABLEAU 18 – Résultats de l'instrument de mesure de la motivation au test  
(CEDRE Compétences langagières et littératie Collège)

<b>Après le test :</b>	<b>Moyenne</b>
Difficulté perçue du test	4,3
Motivation au test	6,9
Motivation au test si les résultats comptaient pour le bulletin scolaire	8,9

FIGURE 22 – Instrument de mesure de la motivation au test

**Question 16**

**Sur une échelle de difficulté allant de 1 à 10, comment avez-vous trouvé les exercices de cette évaluation ?**

Très faciles Très difficiles

<sub>1</sub>   <sub>2</sub>   <sub>3</sub>   <sub>4</sub>   <sub>5</sub>   <sub>6</sub>   <sub>7</sub>   <sub>8</sub>   <sub>9</sub>   <sub>10</sub>   C3HQQ1290201  
247

**Question 17**

**Comment vous êtes-vous appliqué(e) pour faire cette évaluation ?**

(Indiquez votre niveau d'application sur une échelle allant de 1 à 10)

Je ne me suis pas du tout appliqué(e) Je me suis énormément appliqué(e)

<sub>1</sub>   <sub>2</sub>   <sub>3</sub>   <sub>4</sub>   <sub>5</sub>   <sub>6</sub>   <sub>7</sub>   <sub>8</sub>   <sub>9</sub>   <sub>10</sub>   C3HQQ1290401  
248

**Question 18**

**Si les résultats de cette évaluation comptaient pour votre bulletin scolaire, comment vous seriez-vous appliqué(e) ?**

(Indiquez votre niveau d'application sur une échelle allant de 1 à 10)

Je ne me serais pas du tout appliqué(e) Je me serais énormément appliqué(e)

<sub>1</sub>   <sub>2</sub>   <sub>3</sub>   <sub>4</sub>   <sub>5</sub>   <sub>6</sub>   <sub>7</sub>   <sub>8</sub>   <sub>9</sub>   <sub>10</sub>   C3HQQ1290301  
249

## 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.
- Garcia, E., Le Cam, M., & Rocher, T. (2015). Méthodes de sondage utilisées dans les programmes d'évaluation des élèves. *Éducation et Formations*, 85-86, 101-117.
- H. de Monestrol, & H. Rogie. (2021). CEDRE 2015 - 2021 - compétences langagières et littératie en fin de collège : des résultats stables en 2021. *Note d'information*, 22.
- Keskpaik., 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.
- Rousseau, S., & Tardieu, F. (2004). *La macro sas cube d'échantillonnage équilibré. documentation de l'utilisateur*. Paris : INSEE.
- 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.
- Warm, T. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54 n°3, 427-450.

## Liste des tableaux

1	Programme du cycle 4 – Français (B.O.E.N. spécial n° 11 du 26 novembre 2015) . . . . .	6
2	Programme du cycle 4 – Éducation aux médias et à l'information (B.O.E.N. spécial n° 11 du 26 novembre 2015) . . . . .	7
3	Typologie des supports exploités en compréhension de l'écrit et de l'oral dans CEDRE CLL . . . . .	7
4	Cadre de compétences en compréhension de l'écrit et de l'oral de CEDRE CLL . . . . .	8
5	Déroulé de l'évaluation Maitrise de la langue 2021 . . . . .	9
6	Exclusions pour la base de sondage - Compétences langagières et littératie Collège . . . . .	27
7	Répartition dans la base de sondage - Compétences langagières et littératie Collège . . . . .	27
8	Répartition dans l'échantillon - Compétences langagières et littératie Collège . . . . .	28
9	Non-réponse des établissements - Compétences langagières et littératie Collège . . . . .	28
10	Non-réponse des élèves - Compétences langagières et littératie Collège . . . . .	28
11	Comparaison entre les marges de l'échantillon et les marges dans la population - Compétences langagières et littératie Collège . . . . .	30
12	Scores moyens et erreurs standard associées - Compétences langagières et littératie Collège . . . . .	30
13	Répartitions en % dans les groupes de niveaux - Compétences langagières et littératie Collège . . . . .	31
14	Erreurs standards des répartitions en % dans les groupes de niveaux - Compétences langagières et littératie Collège . . . . .	31
15	Effet du plan de sondage - Compétences langagières et littératie Collège . . . . .	32
16	Niveaux de compétences (moyennes des scores et écarts-types) Compétences langagières et littératie Collège . . . . .	50
17	Quel est votre degré de motivation pour faire cette évaluation ? (CEDRE Compétences langagières et littératie Collège) . . . . .	71
18	Résultats de l'instrument de mesure de la motivation au test (CEDRE Compétences langagières et littératie Collège) . . . . .	71

## Table des figures

1	Exemple de liseuse à 3 onglets avec barre de défilement à utiliser sur la page 1. . . . .	11
2	Répartition des items selon l'année de conception et le domaine .	12
3	Répartition des supports en compréhension de l'écrit par type . .	12
4	Répartition des items en compréhension de l'écrit par compétence	13
5	Répartition des items de compréhension de l'oral par compétence	13
6	Répartition des items selon leur format en compréhension de l'écrit et de l'oral . . . . .	14
7	Exemple de format QCM . . . . .	15
8	Exemple de format tableau série . . . . .	16
9	Exemple de format champ libre . . . . .	17
10	Exemple de format glisser-déposer . . . . .	18
11	Représentation graphique utilisée pour le regroupement d'items .	38
12	Modèle de réponse à l'item - 2 paramètres . . . . .	41
13	Exemples d'ajustements (FIT) . . . . .	45
14	Taux de réussite des items en 2015 en fonction de la position de leurs blocs . . . . .	49
15	Principes de construction de l'échelle . . . . .	52
16	Support groupe <1 . . . . .	55
17	Support groupe 1 . . . . .	57
18	Support groupe 2 . . . . .	59
19	Support et question groupe 3 . . . . .	62
20	Support et question groupe 4 . . . . .	65
21	Support et question groupe 5 . . . . .	67
22	Instrument de mesure de la motivation au test . . . . .	72