

CEDRE

Cycle des Évaluations Disciplinaires Réalisées sur Échantillons

Rapport technique

Maîtrise de la langue 2021

École

Auteurs :

Sandra ANDREU

Yann ETEVE

Marguerite GARNERO

Xuân NGHIEM

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

| | |
|---|-----------|
| Introduction | 1 |
| 1 Cadre d'évaluation | 3 |
| 1.1 Objectifs | 3 |
| 1.2 Les compétences et connaissances évaluées | 4 |
| 1.3 Construction du test | 8 |
| 1.4 Passation des évaluations | 29 |
| 2 Sondage | 32 |
| 2.1 Méthodes | 32 |
| 2.2 Échantillonnage | 35 |
| 2.3 État des lieux de la non-réponse | 37 |
| 2.4 Redressement | 40 |
| 2.5 Précision | 40 |
| 3 Analyse des items | 43 |
| 3.1 Comparabilité des évaluations | 43 |
| 3.2 Méthodologie psychométrique | 44 |
| 3.3 Codage des réponses aux items | 47 |
| 3.4 Résultats | 51 |
| 4 Modélisation | 52 |
| 4.1 Méthodologie | 52 |
| 4.2 Résultats | 58 |
| 4.3 Calcul des scores | 60 |
| 5 Construction de l'échelle | 61 |
| 5.1 Méthode | 61 |
| 5.2 Caractérisation des groupes de niveaux | 62 |
| 5.3 Exemples d'items | 65 |
| 6 Variables contextuelles et non cognitives | 74 |
| 6.1 Variables sociodémographiques et indice de position sociale | 74 |
| 6.2 Élaboration des questionnaires de contexte | 75 |
| 6.3 Finalités assignées à la lecture par les élèves en 2021 | 76 |
| 7 Annexe | 80 |
| 7.1 Textes utilisés dans l'évaluation | 82 |
| Références | 90 |

Introduction

Introduction

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

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

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

1 Cadre d'évaluation

1.1 Objectifs

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

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

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

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

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

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

| Discipline évaluée | Début du cycle | Reprises | | |
|--|----------------|----------|------|------|
| Maîtrise de la langue et compétences générales | 2003 | 2009 | 2015 | 2021 |
| Langues étrangères | 2004 | 2010 | 2016 | 2022 |
| Attitude à l'égard de la vie en société | 2005 | – | – | – |
| Histoire, géographie et éducation civique | 2006 | 2012 | 2017 | 2023 |
| Sciences | 2007 | 2013 | 2018 | 2024 |
| Mathématiques | 2008 | 2014 | 2019 | 2025 |

1.2 Les compétences et connaissances évaluées

L'évaluation CEDRE en fin d'école en maîtrise de la langue a pour objectif de faire le point sur les connaissances et les compétences des élèves tant au niveau des savoirs que des savoir-faire. Il s'agira de mesurer l'évolution de ces connaissances et compétences entre 2003, 2009, 2015 et 2020 (reporté en 2021 à cause de la crise sanitaire).

Les connaissances et compétences telles qu'elles sont définies dans le socle commun de connaissances et de compétences ainsi que les programmes officiels constituent le cadre de cette évaluation.

L'élaboration des grilles de compétences pour la construction des items a pris appui sur les documents de référence que sont les programmes officiels en vigueur à partir de la rentrée scolaire 2016-2017 parus au BO spécial n°11 du 26 novembre 2015 avec les ajustements parus au BO n°11 du 26 juillet 2018 (tableau 2).

Dans ce tableau 2 ont été indiqués en gras les éléments évalués dans le CEDRE. C'est sur la base de ces connaissances et compétences du volet d'enseignement du français qu'ont été conçus :

- les items de compréhension de l'écrit et d'étude de la langue pour les évaluations format papier et format numérique (tableau 3 et tableau 4) ;
- et les items de compréhension de l'oral pour l'évaluation format numérique (tableau 5).

TABLEAU 2 – Programme du cycle 3 - enseignement du Français (BOEN n°31 du 30 juillet 2020)

| Thèmes | Attendus de fin de cycle | Compétences et connaissances associées |
|-------------------------------------|--|---|
| Langage oral | Ecouter un récit et manifester sa compréhension en répondant à des questions sans se reporter au texte | Ecouter pour comprendre un message oral, un propos, un discours, un texte lu |
| | Dire de mémoire un texte à haute voix | |
| | Réaliser une courte présentation orale en prenant appui sur des notes ou sur diaporama ou autre outil numérique | Parler en prenant en compte son auditoire |
| Lecture et compréhension de l'écrit | Interagir de façon constructive avec d'autres élèves dans un groupe pour confronter des réactions ou des points de vue | |
| | Lire, comprendre et interpréter un texte littéraire adapté à son âge et réagir à sa lecture | Renforcer la fluidité de lecture |
| | Lire et comprendre des textes et des documents (textes, tableaux, graphiques, schémas, diagrammes, images) pour apprendre dans les différentes disciplines | Comprendre un texte littéraire et l'interpréter Comprendre des textes, des documents et des images et les interpréter |
| Ecriture | | Contrôler sa compréhension et adopter un comportement de lecteur autonome |
| | Ecrire un texte d'une à deux pages adapté à son destinataire. | Ecrire à la main de manière fluide et efficace |
| | Après révision, obtenir un texte organisé et cohérent, à la graphie lisible et respectant les régularités orthographiques étudiées au cours du cycle. | Recourir à l'écriture pour réfléchir et pour apprendre Produire des écrits variés en s'appropriant les différentes dimensions de l'activité d'écriture |
| | | Réécrire à partir de nouvelles consignes ou faire évoluer son texte Prendre en compte les normes de l'écrit pour formuler, transcrire et réviser |
| Etude de la langue | En rédaction de textes dans des contextes variés, maîtriser les accords dans le groupe nominal (déterminant, nom, adjectif), entre le verbe et son sujet dans des cas simples (sujet placé avant le verbe et proche de lui, sujet composé d'un groupe nominal comportant au plus un adjectif ou un complément du nom ou sujet composé de deux noms, sujet inversé suivant le verbe) ainsi que l'accord de l'attribut avec le sujet. | Maîtriser les relations entre l'oral et l'écrit |
| | | Acquérir l'orthographe grammaticale |
| | | Acquérir l'orthographe lexicale |
| | Raisonnement pour analyser le sens des mots en contexte et en prenant appui sur la morphologie. | Enrichir le lexique |
| | Être capable de repérer les principaux constituants d'une phrase simple et complexe. | Identifier les constituants d'une phrase simple, se repérer dans la phrase complexe |

TABLEAU 3 – Compétences évaluées en compréhension de l'écrit dans l'évaluation Maîtrise de la langue 2021 (papier et numérique)

| Intitulé | Compétences | Description |
|------------------|--|---|
| Prélever | Prélever une information explicite, retrouver/localiser une information présente dans le texte. Identifier et mémoriser des informations importantes | Questions littérales : La réponse est prélevée directement dans le texte. Elle peut nécessiter la compréhension de paraphrases. |
| Inférer niveau 1 | Déduire/construire une information en mettant en relation des informations textuelles entre elles. Inférer/Généraliser | Cohérence locale : La réponse peut nécessiter la compréhension d'une causalité implicite, d'une anaphore, d'un cadre spatio-temporel implicite, l'inférence du sens d'un mot non connu. |
| Inférer niveau 2 | Déduire/construire une information en mettant en relation des informations textuelles et des connaissances personnelles. Inférer/Généraliser | Cohérence globale : La réponse peut nécessiter la compréhension d'une inférence de connaissances (question sur les sentiments des personnages) ou la compréhension du thème et/ou des idées principales. |
| Synthétiser | Organiser, résumer, synthétiser | Organiser les idées principales à l'aide d'organiseurs graphiques, choisir un résumé, un titre, ordonner/souligner les idées principales... |

TABLEAU 4 – Compétences évaluées en étude de la langue dans l'évaluation
Maîtrise de la langue 2021 (papier et numérique)

| Intitulé | Compétences | Description |
|-------------|--|--|
| Grammaire | Maîtriser les accords de l'attribut avec son sujet, maîtriser les accords dans le groupe nominal (déterminant, nom, adjectif), identifier les constituants d'une phrase simple, se repérer dans la phrase complexe | <ul style="list-style-type: none"> — Reconnaître le verbe — Repérer les constituants d'une phrase — Accorder le verbe et son sujet |
| Orthographe | Acquérir l'orthographe grammaticale, acquérir l'orthographe lexicale et maîtriser les relations entre l'oral et l'écrit | <ul style="list-style-type: none"> — Accorder déterminant, nom et adjectif — Respecter les correspondances graphèmes-phonèmes — Respecter l'orthographe lexicale des mots les plus courants |
| Conjugaison | Maîtriser les accords entre le verbe et son sujet | <ul style="list-style-type: none"> — Conjuguer les verbes fréquents |
| Vocabulaire | Raisonnement pour analyser le sens des mots en contexte et en prenant appui sur la morphologie. | <ul style="list-style-type: none"> — Reconnaître des mots de la même famille — Reconnaître des synonymes — Utiliser un dictionnaire — Identifier des termes génériques |

TABLEAU 5 – Compétences évaluées en compréhension de l'oral dans l'évaluation
Maîtrise de la langue 2021 (numérique)

| Intitulé | Compétences | Description |
|--------------|--|--|
| Langage oral | Écouter pour comprendre un message oral, un propos, un discours, un texte lu | <ul style="list-style-type: none"> — Mettre en relation des informations textuelles entre elles — Mettre en relation des informations textuelles et des connaissances personnelles — Identifier et mémoriser des informations importantes — Inférer (causalités, personnages) à partir d'un texte oralisé — Choisir un résumé |

1.3 Construction du test

Le cycle complet de cette évaluation s'est déroulé sur 6 années au lieu de 5 en raison de la pandémie de COVID.

TABLEAU 6 – Déroulé de l'évaluation Maîtrise de la langue 2021

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

1.3.1 Conception des items

Les items produits dans une évaluation CEDRE sont le fruit d'un travail collectif des concepteurs, encadrés par le chargé d'étude, personnel du bureau de la conception et du pilotage des évaluations des élèves de la DEPP, sous la responsabilité du chef de bureau. Les concepteurs ont une bonne connaissance des pratiques de classe : ils sont conseillers pédagogiques, professeurs des écoles, maîtres formateurs ou inspecteurs de circonscription. Lors du cadrage de l'évaluation, il est également fait appel à l'expertise de l'Inspection Générale de l'Éducation, du Sport et de la Recherche (IGESR).

Un item proposé par un concepteur fait l'objet d'une discussion contradictoire jusqu'à aboutir à un consensus, au final validé par le chargé d'étude et l'inspection générale. L'item fait alors l'objet d'une expérimentation afin d'être validé sur le plan statistique. S'il est retenu, l'item est intégré à l'évaluation.

Une application ad hoc (cf. plus loin l'encadré « 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 papier

Dans le cadre d'une évaluation sur support papier, le test se compose d'un ensemble de cahiers, constitués de blocs, qui sont eux-mêmes composés d'unités ou situations (ensemble d'items composant un exercice cohérent).

Pour l'évaluation CEDRE Maîtrise de la langue 2021, 49 situations ont été constituées et réparties au sein de 12 cahiers. 11 des 49 situations concernent la compréhension de l'écrit.

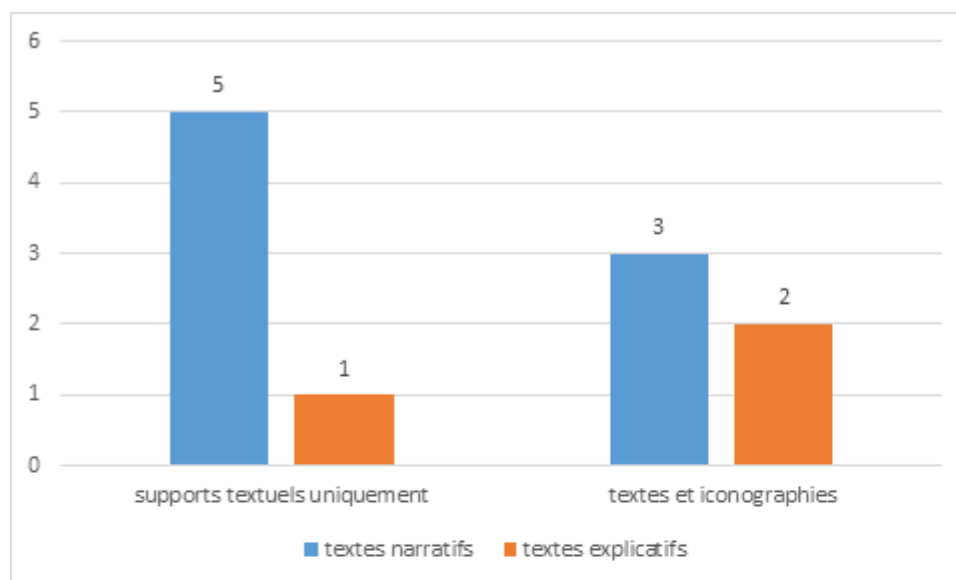
L'évaluation papier CEDRE Maîtrise de la langue 2021 est totalement composée d'items repris à l'identique par rapport à 2003, 2009 et 2015, qu'on appelle « items d'ancrage ».

La compréhension de l'écrit comporte 81 items :

- 31 des items (3 situations) ont été créés entre 2003 et 2009 ;
- 50 des items (8 situations) ont été créés en 2015.

L'évaluation CEDRE Maîtrise de la langue 2021 est composée de différents supports. Sur les 11 situations de compréhension de l'écrit, 8 sont des textes narratifs (soit 72 % des textes) et 3 sont des textes explicatifs. Les supports utilisés présentent soit seulement du texte écrit, soit du texte écrit avec de l'iconographie.

FIGURE 1 – Répartition des types de textes selon le support proposé



Pour rappel, quatre types de questions interrogent ces textes narratifs et explicatifs, avec des items de :

- prélèvement
- inférences de niveau 1
- inférences de niveau 2
- synthèse

FIGURE 2 – Répartition des items de compréhension de l'écrit par compétences

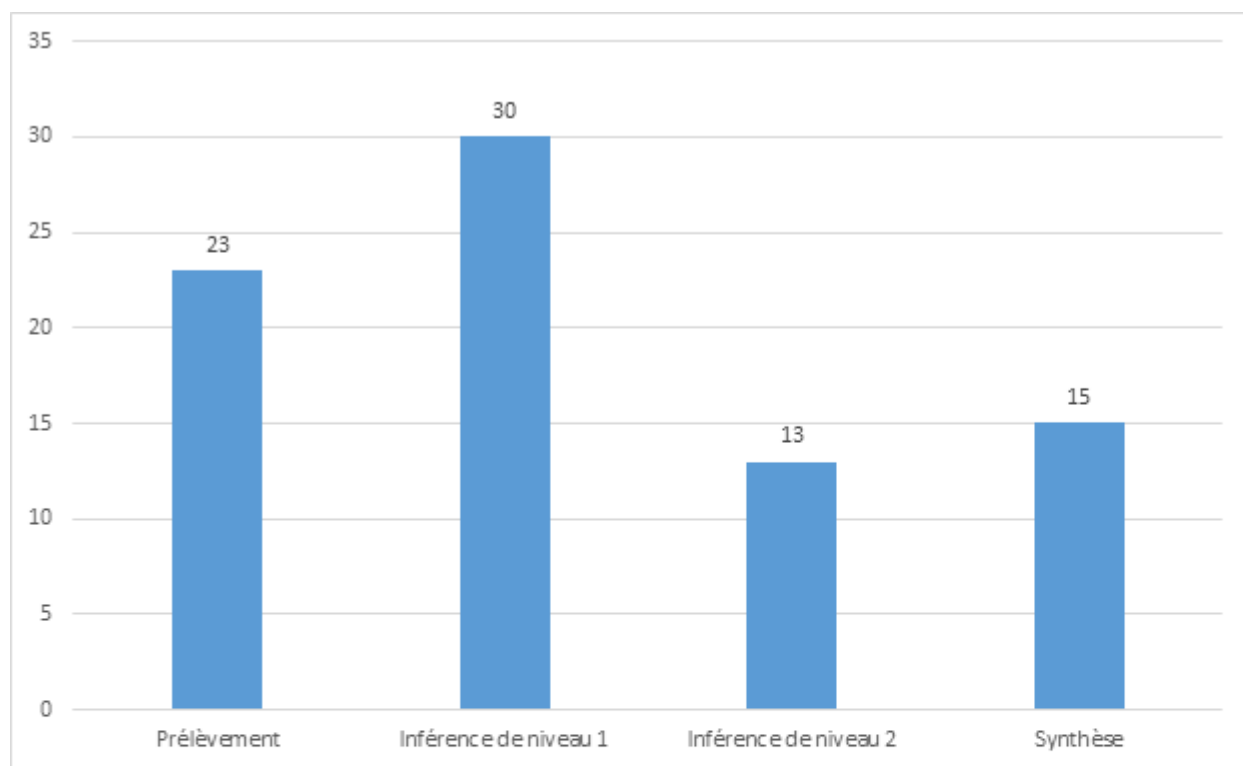
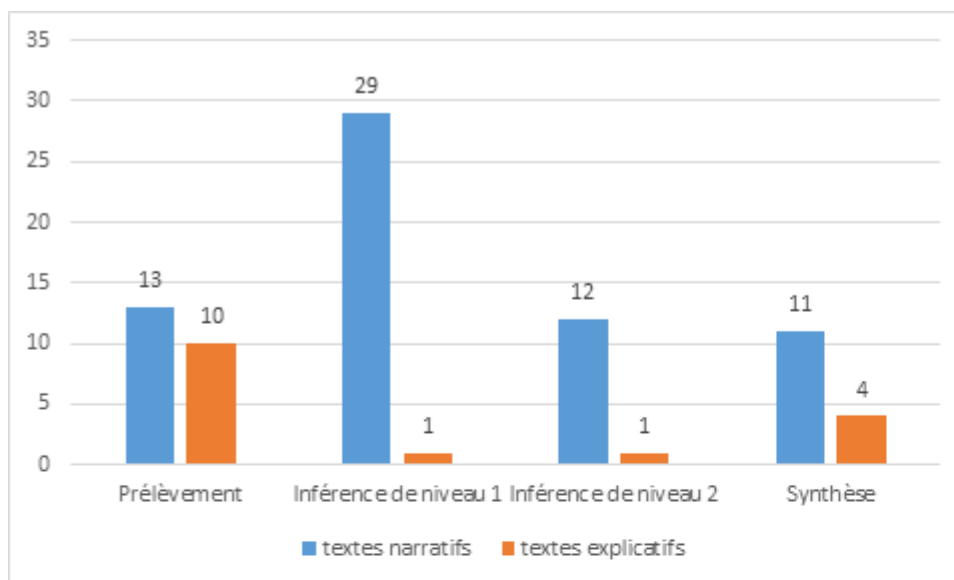
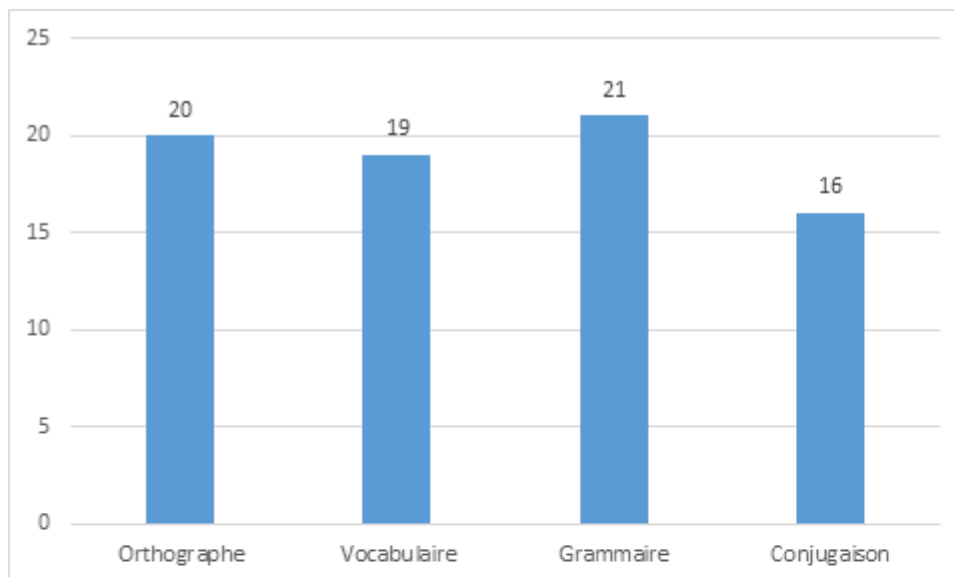


FIGURE 3 – Répartition des types de textes selon les compétences évaluées (en nombre d'items)



La partie Etude de la langue de l'évaluation papier est constituée de 75 items, répartis entre les compétences de grammaire, de vocabulaire, d'orthographe et de conjugaison.

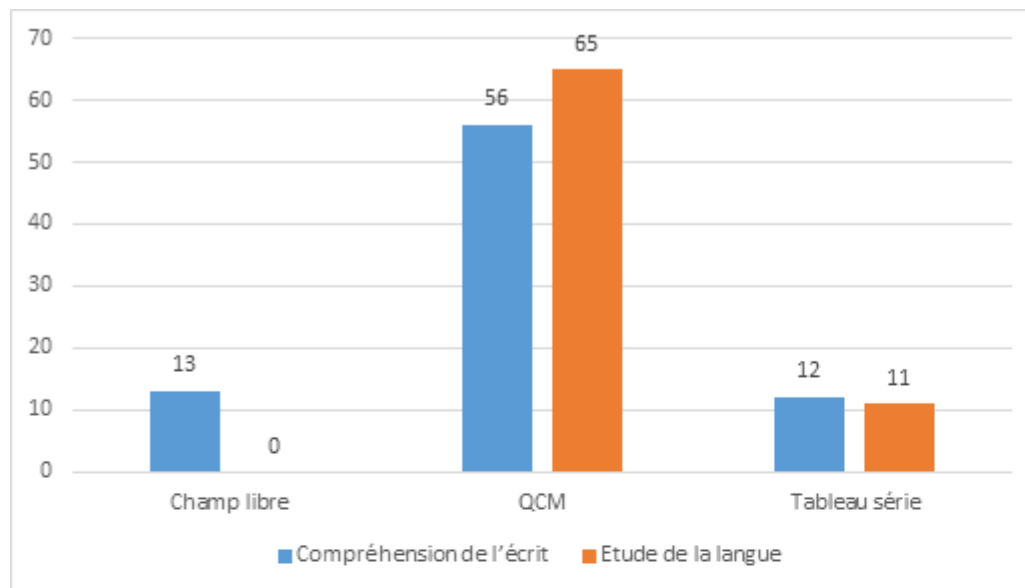
FIGURE 4 – Répartition des items d'étude de la langue par compétences



1.3.3 Formats d'items du test papier

Trois formats de questions ont été principalement utilisés pour l'évaluation papier : QCM, tableau série, champ libre.

FIGURE 5 – Répartition des items d'étude de la langue par compétences



1.3.3.a Le questionnaire à choix multiples (QCM)

Le QCM (questionnaire à choix multiples) a pour avantages 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 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'informations. La question à choix multiples est constituée de trois composantes :

- le prélèvement ;
- la tâche : action effectuée par l'élève pour donner sa réponse. Des préconisations générales sont fournies aux élèves avant de commencer l'évaluation ;
- la consigne, qui pose la question ou définit le problème. La consigne doit être simple et facilement compréhensible ;

- le choix d'une solution : une série d'au moins quatre propositions, parmi lesquelles figurent la bonne réponse et d'autres propositions nommées distracteurs.

FIGURE 6 – Exemple de format QCM

Dès qu'il une odeur de cuisine, ce gourmand arrive.

- 1 sent
 2 sens
 3 sang
 4 sans

1.3.3.b Le tableau série

Le format tableau série permet, pour un même objet, de demander à l'élève de fournir une réponse plus précise. Pour chaque proposition, nous attendons une réponse de type « vrai » ou « faux ». Le score n'est pas calculé sur une seule réponse mais sur l'ensemble des propositions. Nous définissons un seuil pour accorder le crédit complet.

FIGURE 7 – Exemple de format tableau série

Pour chaque phrase, choisis le mot qui convient : cour, court ou cours.

| Question | | | | |
|--|---------------------------------------|---------------------------------------|---------------------------------------|------------------|
| | cour | court | cours | |
| J'ai pris le trajet le plus | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | E5FFG801101 3 |
| Chaque matin, les enfants jouent dans la | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | E5FFG801102 4 |
| Mercredi, ton ... de danse est annulé. | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | E5FFG801103 5 |
| Dépêche-toi, ... plus vite ! | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | E5FFG801104 6 |

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.

FIGURE 8 – Exemple de format champ libre

Question 4

Pourquoi M. Labon sourit-il en voyant qu'il n'y a pas de souris dans les pièges ?

| |
|--|
| |
| |
| |
| |
| |

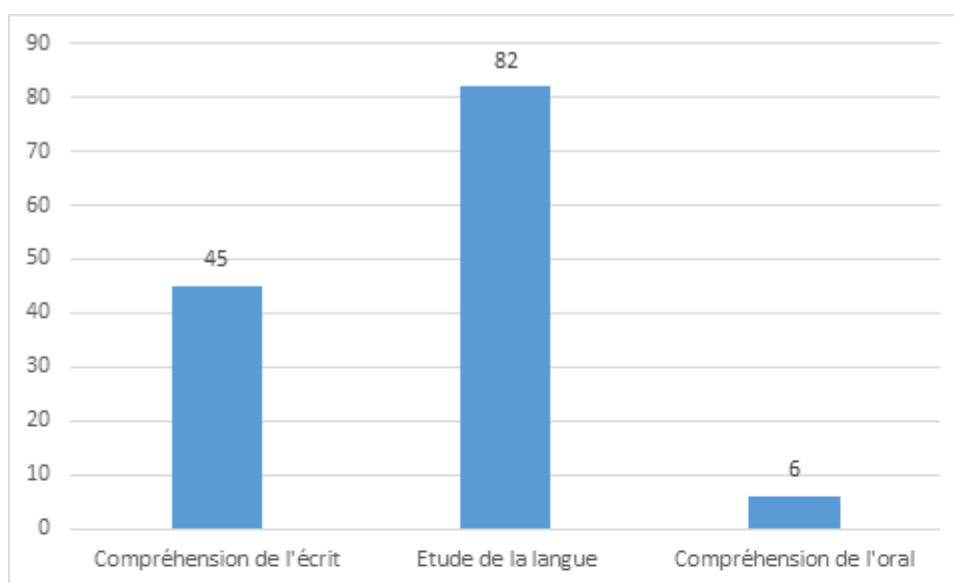
E5FIN1260401
80**1.3.3.d Constitution du test numérique**

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 sous format tablette. L'hétérogénéité de l'équipement informatique dans les écoles suppose que la présence d'ordinateurs en nombre suffisant pour assurer la passation de l'évaluation numérique dans toutes les classes sélectionnées n'est pas assurée. L'utilisation de tablettes qui peuvent être transportées et prêtées d'une école à l'autre et gérées par des administrateurs de test garantit la faisabilité d'une évaluation numérique dans le premier degré.

L'évaluation numérique CEDRE Maîtrise de la langue 2021 comporte au total 133 items retenus sur la base de leur validité statistique après expérimentation. Cela a permis d'élaborer 30 situations réparties dans 12 modules. Ces situations concernent 3 domaines :

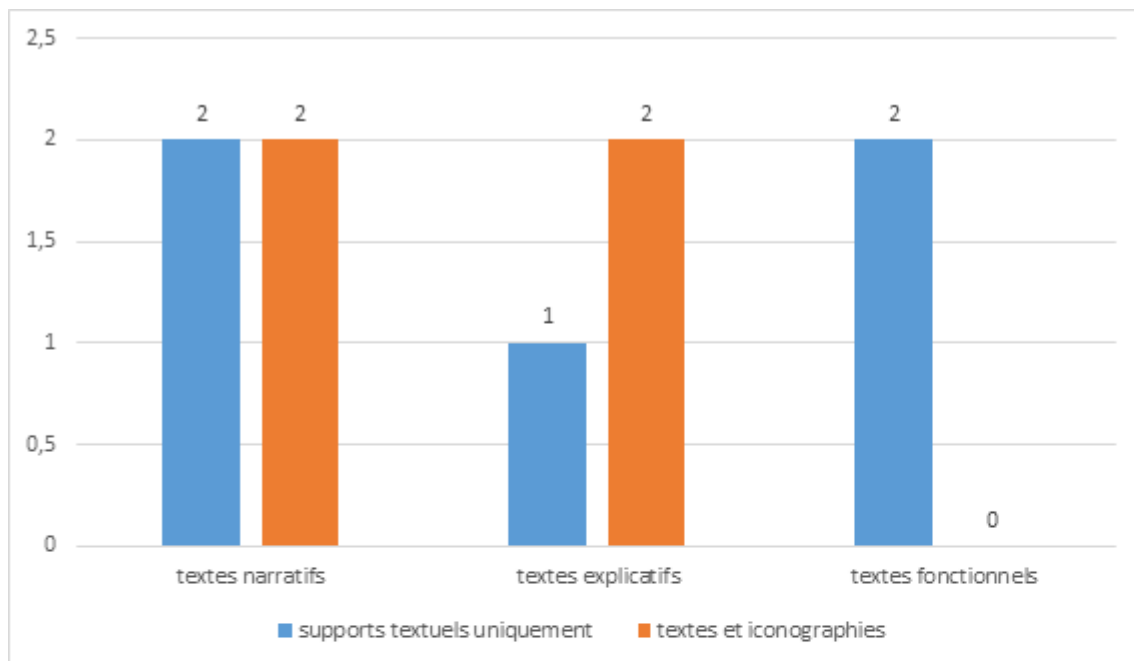
- la compréhension de l'écrit
- l'étude de la langue
- la compréhension de l'oral

FIGURE 9 – Répartition des items numériques par domaine



Les situations en compréhension de l'écrit comportent des textes narratifs, des textes explicatifs, des textes fonctionnels (énoncés de problèmes). Certains présentent seulement du texte, d'autres de l'iconographie.

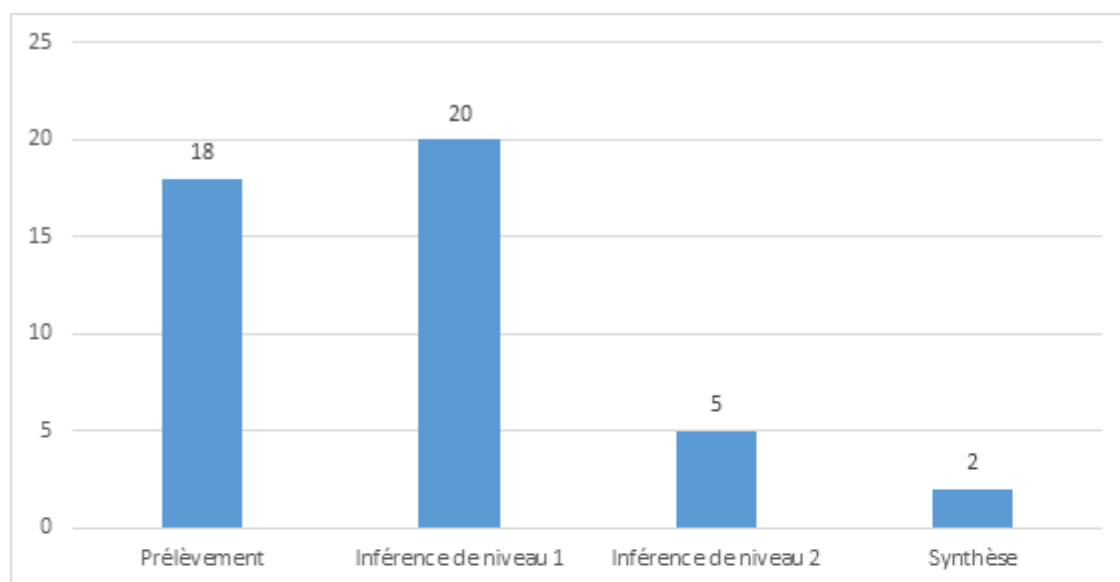
FIGURE 10 – Répartition des types de textes selon le support proposé



Les 45 items de Compréhension de l'écrit évaluent les mêmes compétences que dans l'évaluation papier :

- prélèvement
- inférences de niveau 1
- inférences de niveau 2
- synthèse

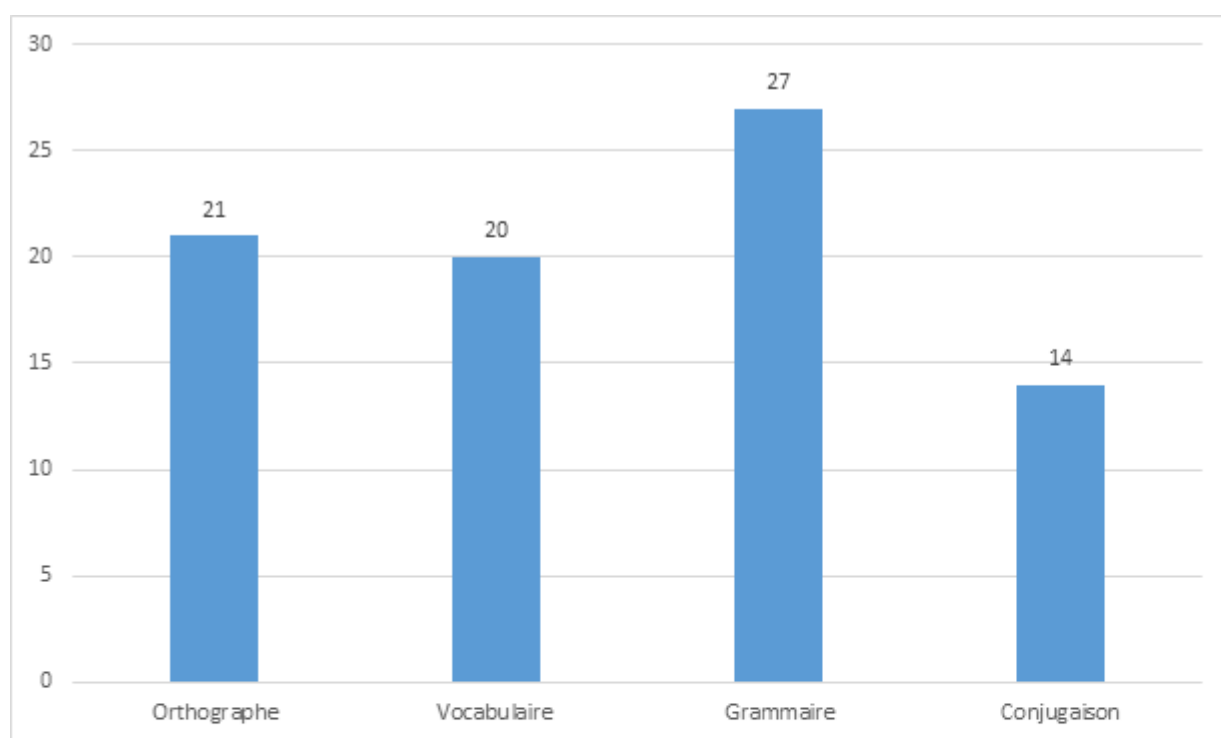
FIGURE 11 – Répartition des items numériques de compréhension de l'écrit par compétences



De même les 82 items d'étude de la langue évaluent les mêmes compétences que l'évaluation papier :

- Orthographe
- Vocabulaire
- Grammaire
- Conjugaison

FIGURE 12 – Répartition des items numériques de d'étude de la langue par compétences



1.3.4 Formats d'items du test numérique

La passation numérique permet la proposition d'un plus grand nombre de formats d'items que l'évaluation papier. Sept formats de questions ont été proposés pour l'évaluation numérique en 2021 :

1.3.4.a Les différents formats de passation numérique

- Phrase cylindre (Figure 13)
- Menu déroulant (Figure 14)

- QCM (Figure 15)
- Surligner (Figure 16)
- Glisser déposer (Figure 17)
- Pointer (Figure 18)
- Champ libre (la réponse ouverte est permise par le biais d'un clavier numérique à faire apparaître). (Figure 19)

FIGURE 13 – Exemple d'item numérique de type phrase cylindre

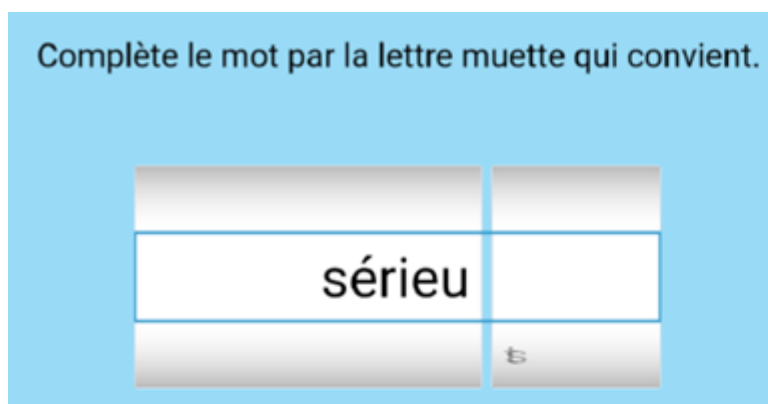


FIGURE 14 – Exemple d'item numérique de type menu déroulant

Choisis la réponse qui convient.

Après la mort du commandant, l'empereur est

heureux.

furieux.

inquiet.

triste.


 Retour au document

FIGURE 15 – Exemple d'item numérique de type QCM

Choisis l'énoncé possible de ce problème mathématique.
Les opérations pour résoudre ce problème sont :

$12 + 3 = 15$
 $20 - 15 = 5$

Louis achète un livre à 12 euros et une trousse à 3 euros. Il paye avec un billet de 20 euros. Combien a-t-il dépensé ?

Louis achète un livre à 12 euros et une trousse à 3 euros. Il paye avec un billet de 20 euros. Combien lui reste-t-il ?

Louis achète un livre et une trousse. Il paye avec un billet de 20 euros. Combien lui reste-t-il ?

Louis achète un livre à 12 euros et une trousse à 3 euros. Il paye avec un billet de 50 euros. Combien lui reste-t-il ?

FIGURE 16 – Exemple d’item numérique de type surligner



FIGURE 17 – Exemple d’item numérique de type glisser déposer

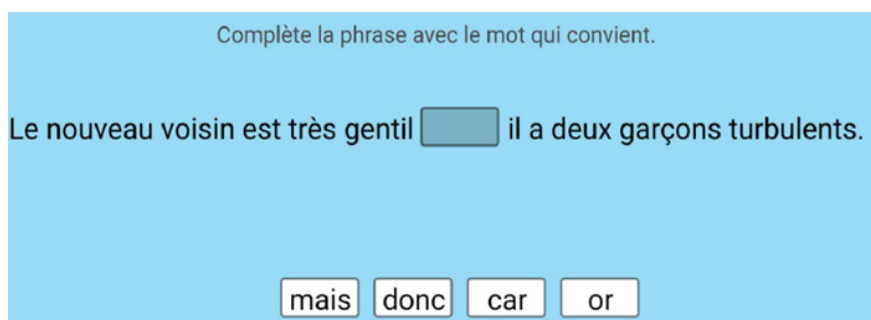


FIGURE 18 – Exemple d'item numérique de type pointer



FIGURE 19 – Exemple d'item numérique de type champ libre

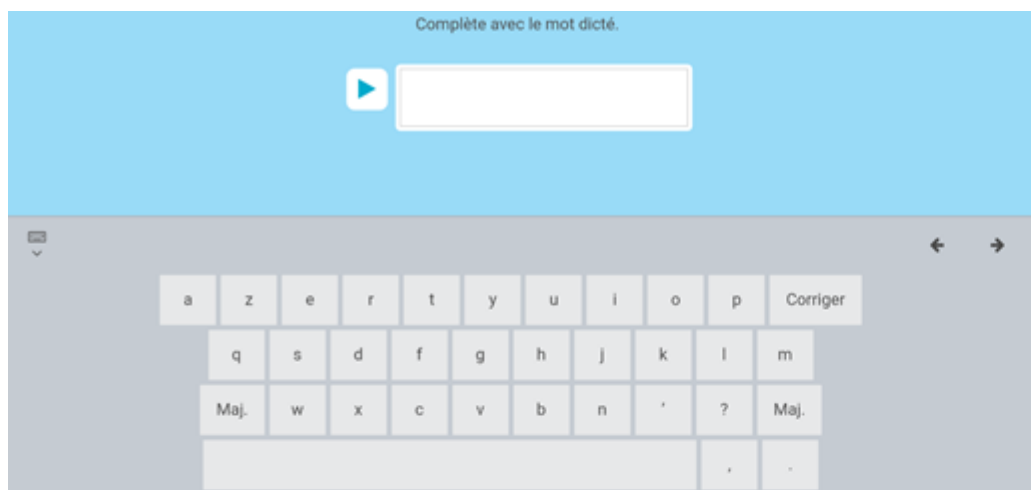
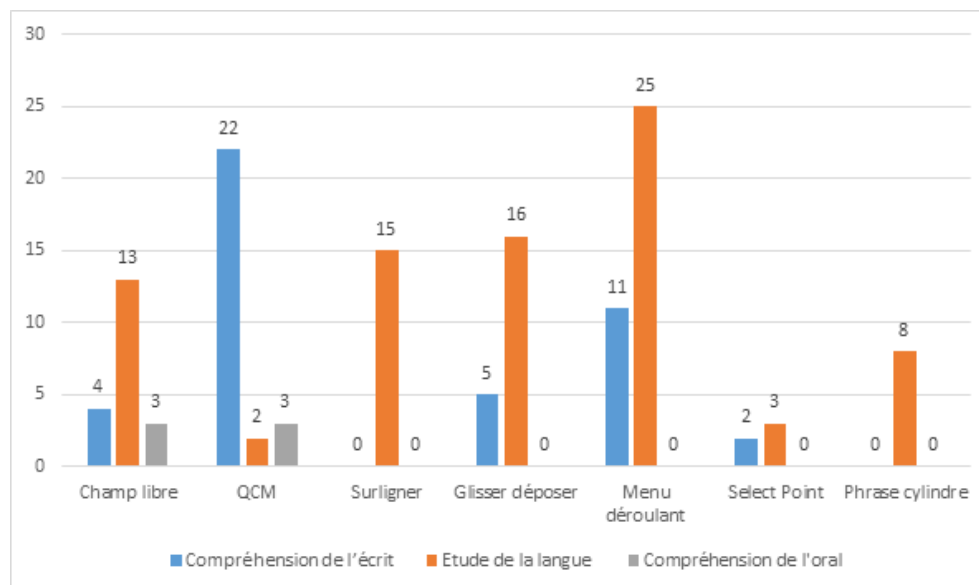


FIGURE 20 – Répartition des items selon leur format



1.3.4.b Utilisation des différents formats numériques selon les dimensions évaluées

Pendant les 5 premières minutes de la passation numérique, les élèves suivent des tutoriels pour découvrir et prendre en main les différentes interactions possibles.

Les items d'étude de la langue utilisent davantage la variété des formats numériques que la compréhension de l'écrit et la compréhension de l'oral. L'avantage de certains gestes est de permettre la mise en avant de la réponse choisie dans le corps de la réponse. Cette ergonomie de la visualisation de la réponse est ainsi un atout en étude de la langue car la vérification de la réponse par l'élève est ainsi facilitée. C'est notamment le cas pour la phrase cylindre, le menu déroulant, le glisser déposer. La fonction de pointage d'un élément est également intéressant sur les supports iconographiques en compréhension de l'écrit.

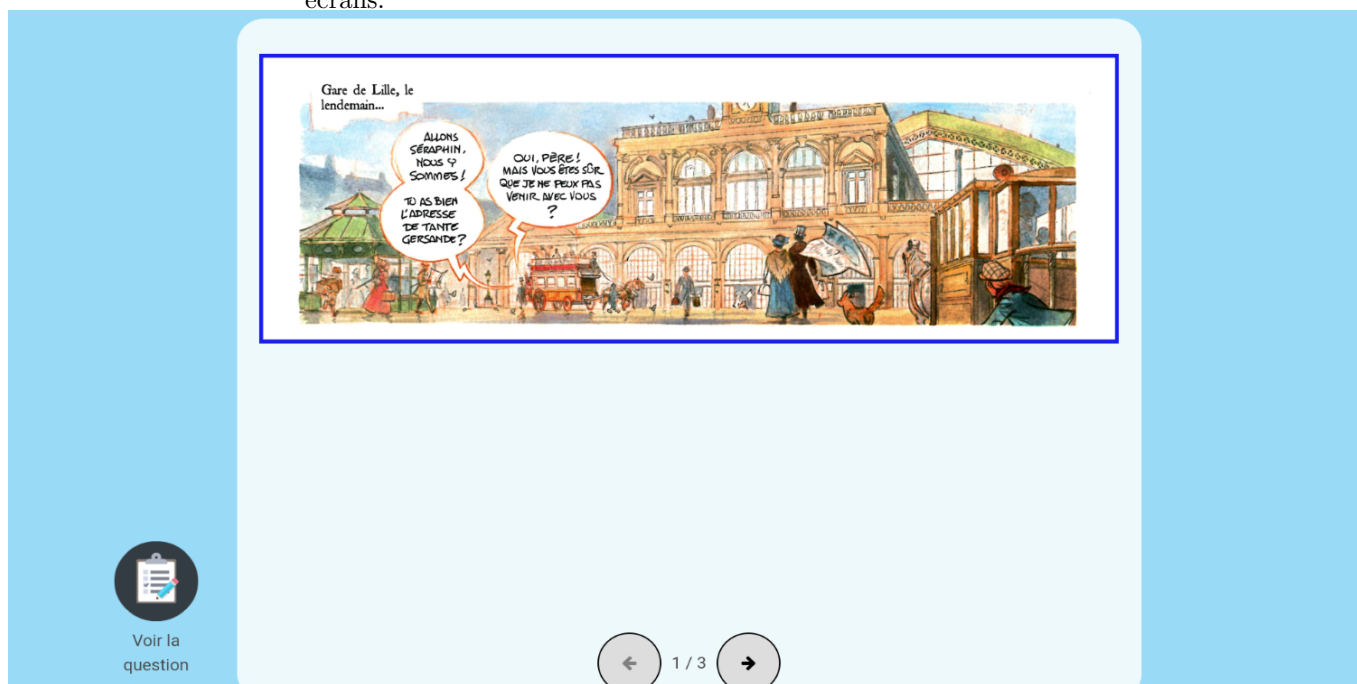
1.3.5 Spécificités des supports sous format numérique

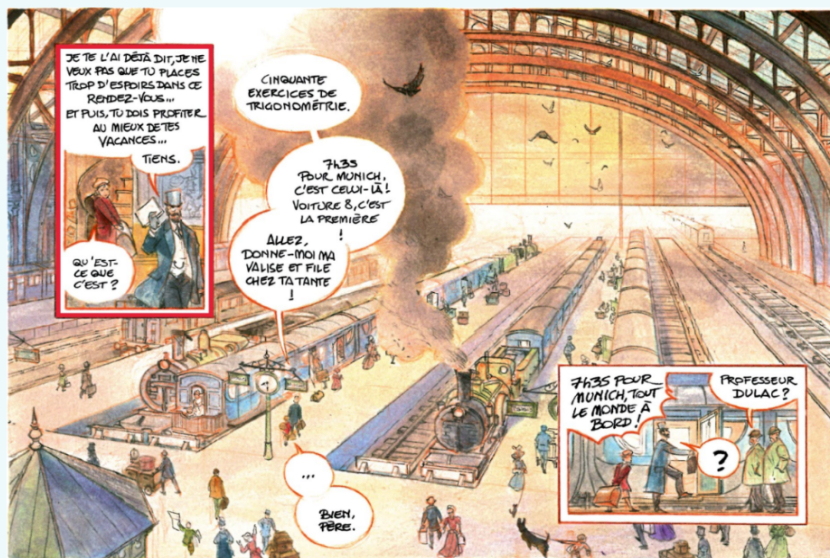
1.3.5.a Lecture de textes sur tablette

La lecture des écrits sur support numérique est différente de celle sur support papier. L'utilisation de supports en couleurs avec une grande définition est rendue possible sur la tablette. L'accès au support est constant avec la possibilité de retourner au document à chaque nouvelle question.

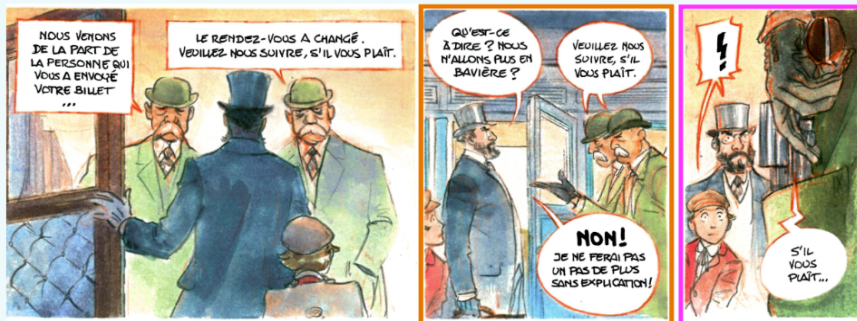
En revanche la taille de l'écran de la tablette peut soulever la question de la visibilité pour les supports excédant la simple page de l'écran. Pour ces derniers, afin d'éviter le scrolling, le choix a été fait de proposer une navigation à l'aide de flèches permettant l'affichage des textes sur des pages successives le cas échéant.

FIGURE 21 – Exemple d'un support de lecture (bande dessinée) réparti sur 3 écrans.





Voir la question



Voir la question



FIGURE 22 – Exemple d'un support de lecture (texte uniquement) réparti sur 2 écrans.

The figure shows two screenshots of a digital reading interface. Each screenshot has a light blue background and a white rounded rectangle containing text. In the bottom-left corner of each screenshot is a circular icon with a clipboard and the text 'Voir la question'. In the bottom-right corner is a navigation bar with left and right arrows and a page indicator.

Screen 1 (Top):

Les pouvoirs de Nasr Eddin.

En Turquie, depuis des siècles, Nasr Eddin Hodja est le héros de nombreuses anecdotes et historiettes amusantes. Il s'oppose dans cette histoire à Timour Leng, un empereur et un grand conquérant qui a vécu au XIV^e siècle.

Il y avait auprès de Timour Leng un homme que Nasr Eddin détestait particulièrement à cause de sa méchanceté : c'était le commandant de sa garde. L'antipathie était réciproque. Un jour qu'il avait fait fouetter un innocent, Nasr Eddin l'aborda et lui annonça pour le troubler :

-Ismail, j'ai consulté les étoiles à ton sujet, et ta mort est prévue pour demain. L'autre ricana et haussa les épaules, mais il se trouva que le lendemain il fit une chute de cheval et se tua sur le coup.

Timour en fut très triste et surtout inquiet car il y vit la preuve que son bouffon Nasr Eddin avait bien, comme on le prétendait, un don de devin et des pouvoirs magiques. Il décida donc de le tuer.

Page indicator: 1 / 2

Screen 2 (Bottom):

Il convoque le Hodja, lequel en entrant dans la salle du trône comprend tout de suite la gravité de la situation : il se trouve encerclé de soldats farouches et le tyran, qui a sa tête des mauvais jours, se tient devant lui, un grand sabre à la main.

-Approche, chien malfaisant. Puisque tu es devin, tu as dû prévoir le jour de ta mort...

-Oui, Seigneur, ce devait être aujourd'hui justement...

-Pourquoi dis-tu cela ? Te prétends-tu assez puissant pour l'éviter ?

-Oui, Seigneur.

-Et comment donc ? fait Timour en levant son arme.

-Très simple, Seigneur : dans les étoiles j'ai vu aussi que ta mort était prévue le lendemain de la mienne.

De colère, Timour jeta son sabre et s'en alla, laissant la vie sauve à son bouffon.

Jean-Louis Maunoury
Les pouvoirs de Nasr Eddin, dans Nasr Eddin Hodja, un drôle d'idiot
Ed. Motus

Page indicator: 2 / 2

1.3.5.b Ecoute de supports oralisés

Le support numérique permet une écoute standardisée des supports de compréhension orale. Il s'agit là d'un atout majeur car les biais éventuels dus à une oralisation de l'enseignant sont ainsi évacués. Le ton, la vitesse de lecture, l'accentuation, la prononciation sont identiques pour tous les élèves en réception. Les élèves disposent tous de casques pour écouter les modules enregistrés. Cela permet l'écoute simultanée de modules différents au sein d'une même classe. Les élèves ont la possibilité de renouveler l'écoute de l'enregistrement s'ils considèrent en avoir besoin. La durée des enregistrements sonores de cette évaluation 2021 varie de 2 secondes (pour une dictée de mots invariables) à 2 minutes (écoute d'une interview de Thomas Pesquet). Ce dispositif permet également l'oralisation des consignes ou des propositions de réponse des exercices concernés.

FIGURE 23 – Exemple d'un item numérique de compréhension de l'oral



1.4 Passation des évaluations

1.4.1 Présentation générale de la passation

L'évaluation s'est déroulée du lundi 31 mai au vendredi 25 juin 2021. Des consignes définies au niveau national et la durée des exercices étaient données dans un guide de passation afin que les conditions de recueil de données et de pas-

sation soient rigoureusement les mêmes dans toutes les écoles concernées.

La passation s'organisait en trois séances. Chaque séquence de cahier était indépendante des autres et faisait l'objet d'une séance en classe. Ces séances pouvaient être organisées dans l'ordre que l'enseignant souhaitait, à raison d'une seule séance par jour.

Les 3 séances se répartissaient ainsi :

- une séance Exercices (séquence 1 du cahier papier) ;
- une séance Dictée / Questionnaire élève (séquence 2 du cahier papier) ;
- une séance Passation numérique sur tablette (prise en charge par un administrateur de test extérieur à l'école).

1.4.1.a L'épreuve papier

Elle était composée de 49 situations réparties dans 12 cahiers différents, d'une dictée et d'un test de fluence. 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.

Les séquences des cahiers étaient organisées ainsi :

- SÉQUENCE 1 : Exercices - 60 minutes
- SÉQUENCE 2 : Dictée - Questionnaires - 30 minutes
- SÉQUENCE 3 : Fluence - 1 minute

L'épreuve de fluence étant individuelle, sa passation se faisait de manière décrochée, dans les délais impartis pour l'évaluation.

Les élèves avaient également à répondre à un questionnaire de contexte qui vise à apporter un éclairage sur les éléments conatifs de l'évaluation. Il se trouvait après la dictée.

1.4.1.b L'épreuve numérique

Pour la passation numérique, 46 situations étaient réparties dans 12 modules différents. Chaque élève passait un module de 45 minutes. Des situations identiques pouvaient se retrouver d'un module à l'autre.

Cette séance était organisée conjointement par l'enseignant de la classe et l'administrateur de test qui encadrait la passation numérique. La passation du test numérique pouvait avoir lieu dans la même salle que la passation papier ou dans une autre salle.

1.4.2 La correction des items

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

- Les questions fermées corrigées de façon automatisées (après numérisation quand elles étaient sous format papier).

- Les questions ouvertes appelant une réponse écrite. Cela concernait des questions de compréhension de l'écrit et la dictée (reprise de celle de 1987 cf NI 22.37 de décembre 2022). Une correction experte est alors nécessaire. Elle est assurée par les concepteurs du groupe Maîtrise de la langue.

Pour les questions ouvertes, un guide de correction décrit les critères de réussite et un codage précis pour chaque type de réponse attendue. Une formation de concepteurs-correcteurs à ce guide permet d'harmoniser les codages et de limiter la subjectivité ou la validation de réponses imprécises. Le codage des réponses aux items d'ancrage, repris d'un cycle d'évaluation à un autre, est strictement identique aux années précédentes.

2 Sondage

2.1 Méthodes

2.1.1 Sondage par grappes stratifié

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

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

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

2.1.2 Redressement de la non réponse : calage sur marges

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

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

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

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

Tirage après élimination de la base des échantillons précédemment tirés

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

Notations

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

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

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

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

Condition fondamentale

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

Champ

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

Sont donc exclus du champ :

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

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

Stratification

La stratification prend en compte le secteur d'enseignement de l'école :

1. écoles publiques hors REP
2. écoles publiques en REP
3. écoles privées

Modalités de sélection

On effectue un tirage d'écoles, stratifié selon trois strates (variable `strate2`). Ensuite, tous les élèves de CM2 des écoles sélectionnées sont interrogés. C'est un sondage par grappes. Les écoles des échantillons des autres opérations de la DEPP ont été préalablement retirées de la base de tirage pour qu'il n'y ait pas de recouvrement.

Échantillon 2021

Pour l'échantillon 2021 (12 cahiers tournants avec 49 situations), on vise 8 000 élèves avec une surreprésentation de l'éducation prioritaire : 3 000 élèves dans

chacune des strates 1 et 2, et 2 000 élèves dans la strate 3, ce qui correspond à 114 écoles dans la strate 1, 83 écoles dans la strate 2 et 70 écoles dans la strate 3.

Base de sondage

Le tableau 7 présente les exclusions dans la population ciblée.

TABLEAU 7 – Exclusions pour la base de sondage - CEDRE Maîtrise de la langue CM2 2021

| | Écoles | Élèves | Écoles exclus | Élèves exclus |
|--|---------------|----------------|------------------|------------------|
| Écoles accueillant des élèves de CM2 | 32 492 | 843 105 | | |
| On retire les TOM | 32 492 | 843 105 | 0 | |
| On retire les écoles hors contrat | 31 825 | 838 018 | 667 | 5 087 |
| On retire les écoles spécialisées | 31 805 | 837 520 | 20 | 498 |
| On retire Mayotte | 316 87 | 830 506 | 118 | 7 014 |
| On retire les petites écoles (<6 CM2) | 29 511 | 822 602 | 2 176 | 7 904 |
| Total exclusions CEDRE | | | 2 981 | 20 503 |
| Base de sondage CEDRE CM2 | 29 511 | 822 602 | | |
| On retire Elaine, PIRLS, CP12, CP/CE1, CP/CE1 num, expé Cedre Allemand, et expé Praesco | 1 524 | 50 023 | | |
| Base de tirage CEDRE MDL CM2 | 27 987 | 772 579 | | |

Le tableau 8 présente la répartition de la population ciblée selon le secteur d'enseignement.

TABLEAU 8 – Répartition base de sondage - CEDRE Maîtrise de la langue CM2 2021

| | Nb d'écoles | Pct d'écoles | Nb d'élèves | Pct d'élèves |
|-------------------|---------------|---------------|----------------|---------------|
| 1. Public hors EP | 21 808 | 73,9% | 574 143 | 69,8% |
| 2. EP | 3 504 | 11,9% | 126 767 | 15,4% |
| 3. Privé | 4 199 | 14,2% | 121 692 | 14,8% |
| Total | 29 511 | 100,0% | 822 602 | 100,0% |

Échantillon

Le tableau 9 présente la répartition de l'échantillon selon le secteur d'enseignement. au total, 518 écoles ont été sélectionnées.

TABLEAU 9 – Répartition dans l'échantillon - CEDRE Maîtrise de la langue CM2 2021

| | Écoles | Élèves |
|-------------------|------------|--------------|
| 1. Public hors EP | 114 | 3 064 |
| 2. EP | 83 | 3 032 |
| 3. Privé | 70 | 2 019 |
| Total | 267 | 8 115 |

2.3 État des lieux de la non-réponse

2.3.1 Non-réponse totale

Parmi la non-réponse totale, nous distinguons selon la non-réponse d'écoles entières ou la non-réponse d'élèves dans les écoles participantes. Les chiffres suivants ont été observés pour 2021. Tout d'abord, 99 % des écoles de l'échantillon ont répondu à l'évaluation (tableau 10).

Au final, 97,2 % des effectifs attendus ont participé (tableau 11).

TABLEAU 10 – Non-réponse des écoles - CEDRE Maîtrise de la langue CM2 2021

| strate | Nb écoles attendues | Nb écoles répondantes | % de écoles répondantes |
|-------------------|---------------------|-----------------------|-------------------------|
| 1- public hors EP | 114 | 114 | 100% |
| 2- EP | 83 | 81 | 97,6% |
| 3- privé | 70 | 69 | 98,6% |
| Total | 267 | 264 | 98,9% |

TABLEAU 11 – Non-réponse globale - CEDRE Maîtrise de la langue CM2 2021

| strate | Nb élèves attendus | Nb élèves répondants | % élèves répondants |
|-------------------|--------------------|----------------------|---------------------|
| 1- public hors EP | 3 064 | 3 045 | 99,4% |
| 2- EP | 3 032 | 2 873 | 94,8% |
| 3- privé | 2 019 | 1 968 | 97,5% |
| Total | 8 115 | 7 886 | 97,2% |

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 retard, afin de pouvoir réaliser des statistiques selon ces variables sur l'échantillon complet, quelle que soit l'analyse. Nous imputons aléatoirement les valeurs manquantes de ces deux variables, de manière à respecter la répartition des répondants.

2.3.3 Non-réponse partielle et terminale

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

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

Dans le premier cas, les non-réponses sont traitées comme des échecs (code "9"). Le second cas conduit à déterminer des règles. Nous considérons que si un élève a passé moins de 50 % d'une séquence, il n'a pas vu la séquence, les valeurs manquantes sont donc traitées de manière structurelle (code "NA").

Pour CEDRE MDL 2021, la non-réponse (partielle et terminale) est étudiée par support (papier et numérique).

Parmi les élèves ayant de la non réponse terminale, il y en a 3.4% pour la séquence papier et 0,4 % pour la séquence numérique.

Si un élève a passé moins de 50 % d'une séquence, on considère qu'il n'a pas vu la séquence (code "s").

Au final, on considère que :

- 231 élèves n'ont pas vu la séquence papier
- 3 élèves n'ont pas vu la séquence numérique

2.4 Redressement

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

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

Le tableau 12 montre que l'ampleur du calage est modérée.

TABLEAU 12 – Comparaison entre les marges de l'échantillon et les marges dans la population

| | Modalité ou variable | Marge échantillon | Marge population | Pourcentage échantillon | Pourcentage population | Pourcentage échantillon après calage |
|--------|-------------------------|----------------------|---------------------|----------------------------|---------------------------|---|
| Retard | 1 | 39 394 | 49 955 | 4,8 | 6,1 | 6,1 |
| | 2 | 783 208 | 772 647 | 95,2 | 93,9 | 93,9 |
| Sexe | 1 | 422 541 | 419 243 | 51,4 | 51 | 51 |
| | 2 | 400 061 | 403 359 | 48,6 | 49 | 49 |
| Strate | 1 | 574 143 | 574 143 | 69,8 | 69,8 | 69,8 |
| | 2 | 126 767 | 126 767 | 15,4 | 15,4 | 15,4 |
| | 3 | 121 692 | 121 692 | 14,8 | 14,8 | 14,8 |

2.5 Précision

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

TABLEAU 13 – Scores moyens et erreurs standard associées - CEDRE Maîtrise de la langue CM2

| Année | Score moyen | Erreur standard |
|-------|-------------|-----------------|
| 2003 | 250 | 1.37 |
| 2009 | 251.0 | 1.45 |
| 2015 | 250.7 | 1.24 |
| 2021 | 254.3 | 1.26 |

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

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

La valeur obtenue est de 2,03 (donc supérieure à 1,96), l'augmentation du score moyen observée entre 2015 et 2021 est donc statistiquement significative, à la différence des évolutions de scores observées entre 2003 et 2009 (0,50), et entre 2009 et 2015 (0,15).

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

TABLEAU 14 – Répartition en % dans les groupes de niveaux - CEDRE Maîtrise de la langue CM2

| Année | Groupe < 1 | Groupe 1 | Groupe 2 | Groupe 3 | Groupe 4 | Groupe 5 |
|-------|------------|----------|----------|----------|----------|----------|
| 2003 | 2.7 | 12.3 | 26.3 | 29.4 | 19.3 | 10.0 |
| 2009 | 2.6 | 11.3 | 25.6 | 31.7 | 19.5 | 9.4 |
| 2015 | 1.0 | 10.0 | 29.1 | 33.1 | 19.7 | 7.2 |
| 2021 | 1.2 | 9.8 | 26.6 | 32.1 | 21.2 | 9.2 |

TABLEAU 15 – Erreurs standards des répartitions en % dans les groupes de niveaux - CEDRE Maîtrise de la langue CM2

| Année | Groupe < 1 | Groupe 1 | Groupe 2 | Groupe 3 | Groupe 4 | Groupe 5 |
|-------|------------|----------|----------|----------|----------|----------|
| 2003 | 0.3 | 0.7 | 0.7 | 0.6 | 0.6 | 0.6 |
| 2009 | 0.3 | 0.7 | 0.7 | 0.8 | 0.9 | 0.6 |
| 2015 | 0.1 | 0.6 | 0.9 | 0.7 | 0.7 | 0.6 |
| 2021 | 0.1 | 0.5 | 0.9 | 0.7 | 0.7 | 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 :

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

Dans le cas d'un sondage en grappes, la précision est dégradée en comparaison d'un sondage aléatoire simple (tableau 16). Cela signifie qu'en 2021, un sondage aléatoire simple avec un effectif 1.75 fois moins important aurait conduit au même niveau de précision.

TABLEAU 16 – Effet du plan de sondage - CEDRE Maîtrise de la langue CM2 2015

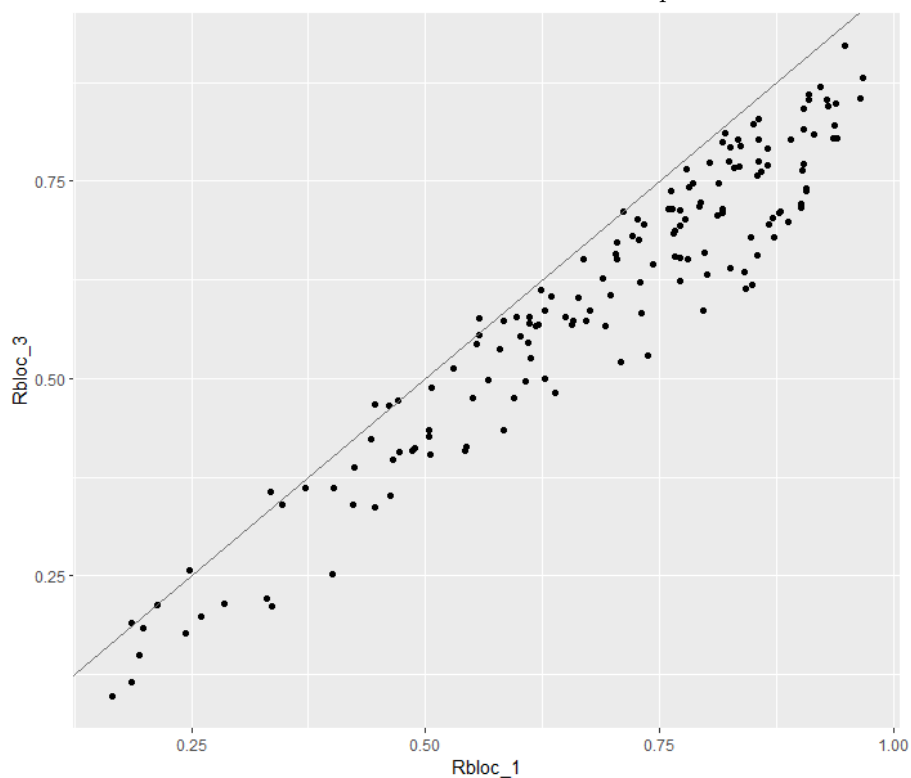
| Année | Erreur Standard | Erreur SAS | <i>Design Effect</i> |
|-------|-----------------|------------|----------------------|
| 2003 | 1.37 | 0.64 | 2.15 |
| 2009 | 1.45 | 0.69 | 2.11 |
| 2015 | 1.20 | 0.50 | 2.41 |
| 2021 | 1.26 | 0.72 | 1.75 |

3 Analyse des items

3.1 Comparabilité des évaluations

Les items de la séquence papier de l'évaluation CEDRE Maîtrise de la langue 2021 étaient répartis en trois blocs d'une durée de 20 minutes qui étaient ventilés dans les 12 cahiers. Les blocs pouvaient donc être passés en début, au milieu ou à la fin d'un cahier. Cependant la durée de l'évaluation 2021 pour cette séquence était 20 minutes plus longues qu'en 2015 ce qui peut conduire à un « Effet fatigue », les élèves seraient moins performants sur les items situés en fin de cahier que sur ceux placés en début de cahier. Pour vérifier cela nous avons comparé les taux de réussite des items en fonction de la position de leurs blocs (figure 24).

FIGURE 24 – Taux de réussite des items en fonction de la position de leurs blocs



On observe que sur cette figure les items sont presque tous mieux réussis lorsqu'ils sont passés en début de cahier, un « Effet fatigue » est bien présent. Dans un souci de comparabilité avec l'évaluation de 2015 nous avons donc décidé de ne garder pour chaque élève que les deux premiers blocs.

3.2 Méthodologie psychométrique

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).

3.2.1 Approche classique

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

Réussite et score

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

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

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

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

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

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

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

La théorie classique des tests a précisément pour objet d'étude le score S_i obtenu par un élève à un test. Elle postule notamment que ce score observé résulte de la somme d'un score « vrai » inobservé et d'une erreur de mesure. Un certain nombre d'hypothèses portent alors sur le terme d'erreur (pour plus d'informations, cf. par exemple Laveault et Gregoire, 2002).

Fidélité

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

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

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

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

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

Indices de discrimination

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

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

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

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

En effet, on peut montrer que

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

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

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

3.2.2 Analyse factorielle des items

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

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

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

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

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

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

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

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

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

3.3 Codage des réponses aux items

3.3.1 Valeurs manquantes

Trois types de valeurs manquantes sont distinguées :

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

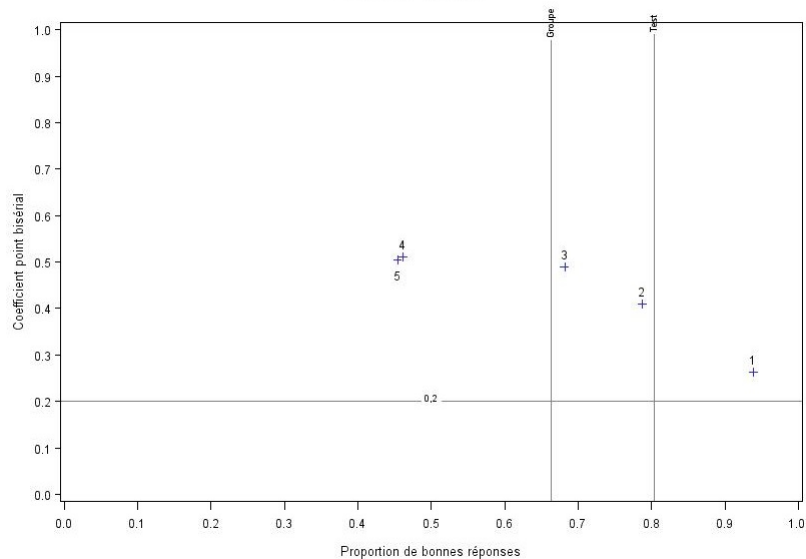
- Non-réponse terminale : l'élève s'est arrêté au cours de l'épreuve, potentiellement en raison d'un manque de temps. Des choix sont effectués pour déterminer le traitement de ces valeurs. Nous considérons que si un élève a passé moins de 50 % d'une séquence, il n'a pas vu la séquence, les valeurs manquantes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

3.3.2 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 25). 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 25 – Représentation graphique utilisée pour le regroupement d'items



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

3.3.3 Traitement des données et correction des questions ouvertes

Tous les cahiers recueillis dans le cadre de cette opération ont été scannés par un prestataire extérieure (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. En ce qui concerne les questions ouvertes, demandant une rédaction plus ou moins longue de la part des élèves (explication, schématisation...), elles ont été découpées en « imassettes » 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 le bureau des évaluations de la DEPP. Il a permis d'intégrer des questions ouvertes dans des évaluations à grandes échelles, aussi bien aux évaluations nationales qu'aux évaluations internationales telles PISA, TIMSS ou PIRLS. Les correcteurs n'ont plus à manipuler un nombre très important de cahiers et peuvent travailler de manière autonome lorsqu'ils le souhaitent, tout en maintenant un contact entre eux et les responsables de l'évaluation afin d'assurer une meilleure fiabilité de la correction.

Principes fonctionnels

Le chef de projet paramètre la session de correction. Il définit les groupes de correcteurs et supervise chaque groupe. Il intègre et vérifie les items mis en correction et ajuste les paramètres de double correction. Son rôle consiste également à répondre aux questions des correcteurs par le biais d'une messagerie intégrée au logiciel et à communiquer sa réponse également aux autres correcteurs. Le superviseur gère son groupe de correcteurs. Il anime la session de formation, qui consiste d'une part à communiquer aux télécorrecteurs une grille de correction très précise et d'autre part à corriger collectivement à blanc un nombre défini d'images 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.

3.4 Résultats

3.4.1 Pouvoir discriminant des items

Le calcul des indices de discrimination nous a conduit à éliminer 13 items dont les indices *rbis-point* étaient trop faibles : 7 items de 2015 et 6 items de 2021.

4 Modélisation

4.1 Méthodologie

4.1.1 Modèle de réponse à l'item

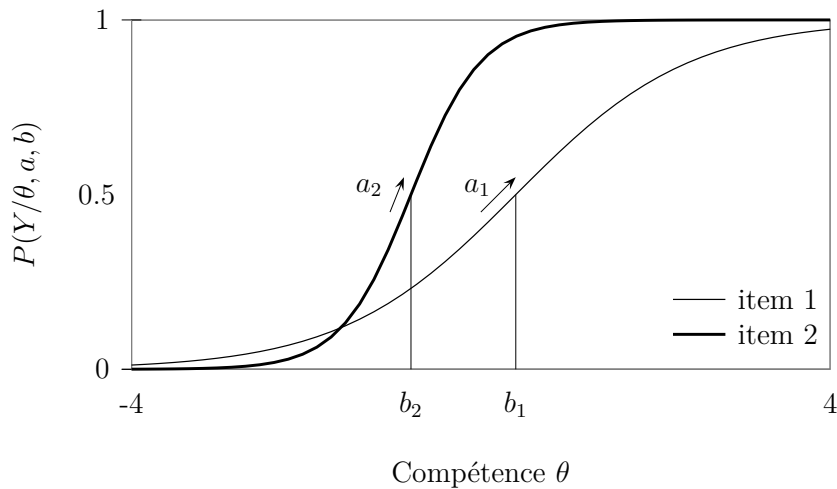
Le modèle de mesure utilisé est un modèle de réponse à l'item à deux paramètres avec une fonction de lien logistique (MRI 2PL) :

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

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

La figure 26 représente les courbes caractéristiques de deux items selon cette modélisation.

FIGURE 26 – 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 26).

4.1.2 Procédures d'estimation

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

Estimation des paramètres des items

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

Notons tout d'abord que les modèles présentés ne sont pas identifiables. En effet, les transformations $\theta_i^* = A\theta_i + B$, $b_j^* = Ab_j + B$ et $a_j^* = a_j/A$ avec A et B deux constantes ($A > 0$), conduisent aux mêmes valeurs des probabilités. Dans CEDRE, nous levons l'indétermination en standardisant la distribution des θ pour les données du premier cycle (en l'occurrence, moyenne de 250 et écart-type de 50 pour l'année 2003).

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

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

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

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

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

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

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

avec g la densité de θ .

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

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

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

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

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

Estimation des niveaux de compétence

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

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

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

avec

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

et

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

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

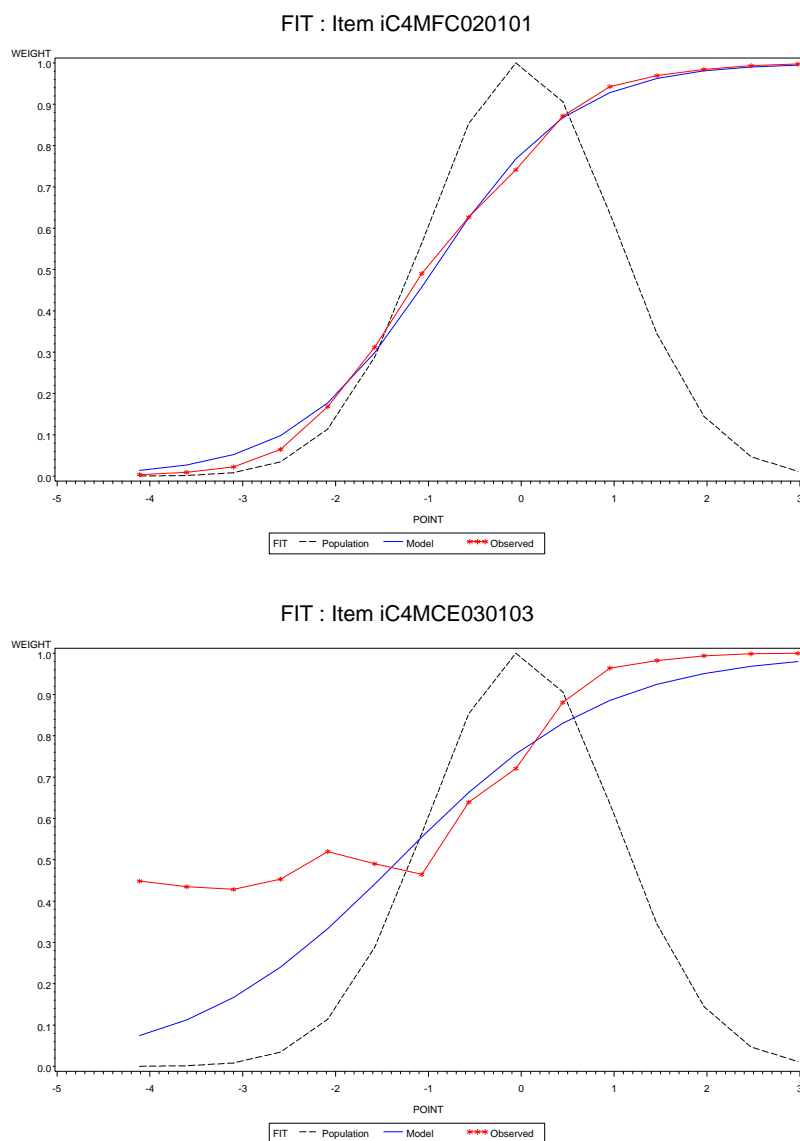
$$\frac{\partial \ln L}{\partial \theta_i} + \frac{J}{2I} = 0 \quad (15)$$

4.1.3 Indice d'ajustement (FIT)

L'ajustement des items au modèle est étudié. Graphiquement, cela revient à comparer les courbes caractéristiques estimées avec les résultats observés (cf. figure 27). Certaines procédures proposent de comparer directement les probabilités théorique avec les proportions de réussite de groupes d'élèves. Plus généralement, nous pouvons écrire les résidus de la manière suivante :

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

FIGURE 27 – Exemples d’ajustements (FIT)



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

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

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

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

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

4.1.4 Fonctionnement Différentiel d'Item (FDI)

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

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

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

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

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

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

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

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

4.1.5 L'information du test

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

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

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

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

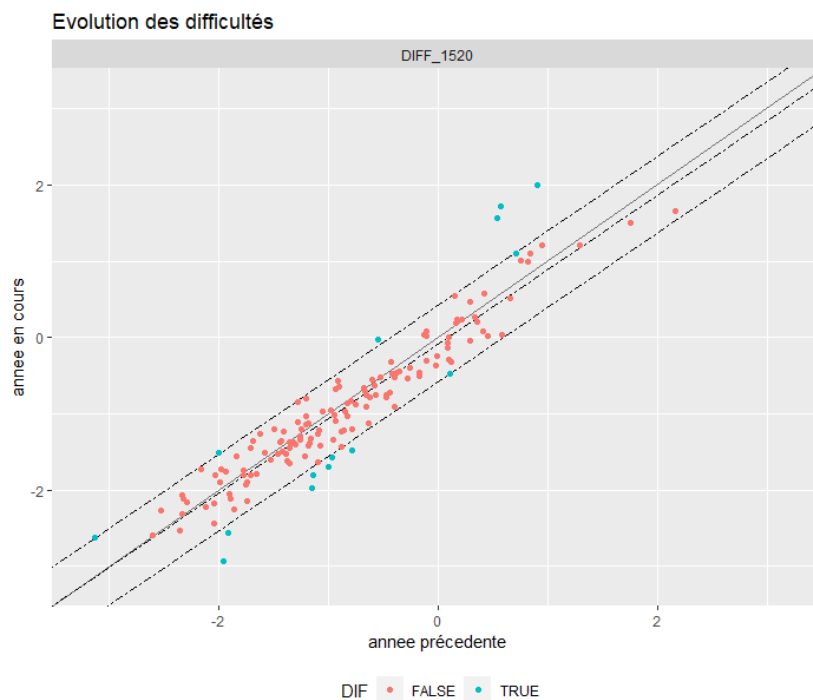
4.2 Résultats

4.2.1 Identification des fonctionnements différentiels d'items (FDI)

L'analyse des FDI 2015-2021 a permis de détecter 17 items en fonctionnement différentiel (figure 28).

Ces items ont été considérés comme de nouveaux items.

FIGURE 28 – Comparaison des paramètres de difficulté 2015-2021 (CEDRE Maîtrise de la langue CM2)



Note de lecture : Les points sont les items. En abscisse figure la valeur des paramètres de difficulté estimés en 2015, et en ordonnée la la valeur des paramètres de difficulté estimés et ajustés pour l'année 2021. Les items présentant un FDI apparaissent en bleu.

4.2.2 Bilan de l'analyse des items

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

- 211 items en 2003
- 171 items en 2009, dont 80 repris de 2003
- 233 items en 2015, dont 85 repris de 2009
- 157 items papiers en 2021, tous présents en 2015
- 219 items numériques en 2021

Cela représente 991 items passés par les élèves en tout, dont 376 en 2021.

Après suppression des items présentant un mauvais Rbis il reste :

- 211 items en 2003
- 171 items en 2009
- 233 items en 2015

- 144 items papiers en 2021
- 219 items numériques en 2021

Une calibration concurrente des données 2021 est réalisée sur les données 2015.

377 items sont donc utilisés dans le cadre de cette calibration concurrente, dont 144 passés dans l'évaluation 2021. En effet la partie numérique de cette opération a surtout une visée exploratoire et n'a donc pas été prise en compte dans l'analyse finale et le calcul des scores.

Pour plus d'informations sur le numérique le lecteur est invité à consulter la Note d'Information (Eteve, Nghiem, & Philbert, 2021).

4.3 Calcul des scores

Comme indiqué précédemment, une analyse conjointe des données (2015 et 2021) a permis d'estimer les paramètres des items, puis les niveaux de compétences θ des élèves pour ces deux années. Les paramètres et niveaux de compétences ainsi calculés ont ensuite été remis sur l'échelle établie en 2003. Afin de lever l'indétermination du modèle, la moyenne des θ a été fixée à 250 et leur écart-type à 50 pour l'échantillon de 2003. Le tableau 17 présente les résultats obtenus.

TABLEAU 17 – Niveaux de compétences CEDRE Maîtrise de la langue École (moyennes et écarts-type)

| annee | Nb élèves | Moyenne | Ecart-Type |
|-------|-----------|---------|------------|
| 2003 | 6 110 | 250.0 | 50.0 |
| 2009 | 4 942 | 251.0 | 48.4 |
| 2015 | 7 428 | 250.7 | 43.0 |
| 2021 | 8 115 | 254.3 | 45.9 |

5 Construction de l'échelle

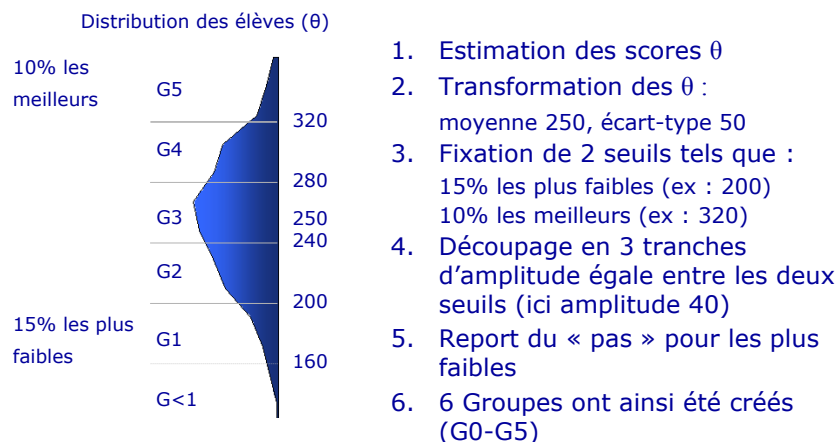
5.1 Méthode

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 Maîtrise de la langue 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 2003. Puis, comme le montre la figure 29, 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 compétences.

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 29 – Principes de construction de l'échelle



5.2 Caractérisation des groupes de niveaux

La mesure fine des acquis des élèves en Maîtrise de la langue, 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° 22.28 (Y. Eteve, X. Nghiem, L. Philbert, 2022).

Groupe < 1 (1,2 % des élèves)

Les élèves de ce groupe sont en très grande difficulté de lecture et en étude de la langue.

La réussite aux exercices de compréhension de l'écrit et d'étude de la langue est difficile.

Groupe 1 (9,8 % des élèves)

Les élèves de ce groupe sont de faibles « compreneurs ». L'élève le plus faible de ce groupe a un score de 160.

Leur compréhension reste locale. Ils ne sont pas en mesure de dresser un portrait global du texte. Ils peuvent mettre en lien des éléments textuels et des images spatialement proches l'un de l'autre et placés en début de document (bande dessinée) pour effectuer des inférences. En étude de la langue, les réussites se limitent au vocabulaire. Ces élèves s'engagent dans une réflexion sur l'utilisation

des mots (repérage des connecteurs logiques pertinents au niveau de la phrase et identification d'une famille de mots).

Groupe 2 (26,6 % des élèves)

Les élèves de ce groupe sont capables de comprendre des écrits plus longs que ceux des groupes inférieurs. L'élève le plus faible de ce groupe a un score de 199.

Ils sont capables de prélever des informations et de faire des inférences dans des écrits plus longs. Ils maîtrisent des compétences relevant principalement du cycle 2 : identifier une reprise anaphorique lorsque le personnage référent est proche du substitut, replacer des événements dans l'ordre chronologique. Ils peuvent mettre en lien deux informations spatialement proches l'une de l'autre pour identifier un lien de causalité. Ils commencent à savoir catégoriser les types d'écrit. Ils identifient des relations logiques dans des textes plus longs, ainsi que les intentions des personnages.

Ils sont capables d'accorder un verbe à un temps simple avec son sujet lorsque celui-ci est à la troisième personne du singulier ou du pluriel. Ils commencent à respecter la concordance des temps lorsque plusieurs verbes conjugués sont présents dans une même phrase. Ils commencent à réaliser les accords dans le groupe nominal. En vocabulaire, ils peuvent trouver le sens figuré d'une expression sans contextualisation et identifier un intrus dans une liste de mots de la même famille.

Ils accèdent à la notion d'homophonie. Ils sont capables de choisir un connecteur logique pour assurer une relation sémantique.

Groupe 3 (32,1 % des élèves)

Les élèves de ce groupe sont des lecteurs capables d'accéder à la compréhension globale d'un écrit long et d'affiner leur appréhension des personnages et du narrateur. L'élève le plus faible de ce groupe a un score de 238.

Ces élèves montrent une bonne compréhension du sens global en construisant des inférences multiples à partir d'informations dispersées dans un texte narratif. Ils effectuent des liens de cause à effet implicites. Ils maîtrisent les notions de personnages et de narrateur, même quand le narrateur est un des personnages. Ils savent synthétiser le sens général d'un texte narratif ou composite à travers le choix d'un titre. Ils sont capables de généraliser le thème d'un texte pour en extraire une morale.

Ils accordent le verbe à un groupe sujet éloigné dans le cas de phrases complexes, ils réalisent des accords en genre et en nombre dans le groupe nominal avec des adjectifs mis en apposition. La maîtrise des homophones grammaticaux se

consolide.

Ils accèdent à la compréhension d'un mot grâce au contexte. Ils peuvent traiter une expression de sens figuré, trouver un synonyme.

Ils ont acquis une maîtrise satisfaisante des compétences et connaissances exigibles en fin d'école.

Groupe 4 (21,2 % des élèves)

Les élèves de ce groupe sont des lecteurs qui déploient des stratégies expertes pour atteindre une représentation globale du texte. L'élève le plus faible de ce groupe a un score de 276.

Ces élèves comprennent les intentions de l'auteur et prélèvent des informations dispersées dans les textes pour formuler des réponses construites à partir d'inférences multiples. Ils effectuent des inférences sur l'état mental des personnages tout en faisant appel à des connaissances personnelles. Ils se représentent mentalement l'histoire et mobilisent leur mémoire de travail sur un texte long. Ils maîtrisent la lecture d'un texte composite : ils effectuent des prélèvements d'indices et en repèrent la fonction.

Ils respectent l'accord sujet-verbe lorsque le sujet est éloigné du verbe. Ils respectent l'accord du participe passé à la voix passive. Ils repèrent l'accord de l'adjectif attribut du sujet. Ils accèdent à la polysémie d'un mot grâce au contexte. Ils maîtrisent une très grande partie des compétences et connaissances en maîtrise de la langue exigibles en fin d'école.

Groupe 5 (9,2 % des élèves)

Les élèves de ce groupe sont des lecteurs critiques qui ont une compréhension de l'évolution de la trame narrative. L'élève le plus faible de ce groupe a un score de 315.

Ces élèves sont capables de saisir l'atmosphère et le ton d'un texte pour comprendre les intentions des personnages. Ils effectuent des inférences sur la personnalité des protagonistes en convoquant des connaissances personnelles. Ils savent interpréter l'évolution des sentiments d'un personnage au fil d'un texte. Ils sont capables d'assurer avec méthode et précision la recherche d'un mot dans un dictionnaire. Ces élèves ont une maîtrise parfaite des compétences et connaissances exigibles en maîtrise de la langue à la fin d'école.

À chaque groupe d'élèves, nous pouvons associer des items de difficulté plus ou moins importante. Si l'item est associé au groupe d'élèves, cela signifie que les élèves de ce groupe ont une probabilité d'au moins 50 % de répondre correctement à cet item. Suivent un ensemble d'items représentatifs de chaque groupe.

5.3 Exemples d'items

5.3.1 Item caractéristique du groupe <1 et 1

Les groupes < 1 et 1 sont constitués des élèves les plus faibles qui représentent 11% des élèves évalués. Ils sont capables de mettre en relation des éléments textuels et des images placés en début de document et spatialement proches l'un de l'autre pour effectuer des inférences. Dans l'exemple qui suit, la compétence visée met en jeu des éléments proches dans le texte pour en dégager le sens.

FIGURE 30 – Exemple support groupe <1 et 1



Support

Extrait d'une planche de bande dessinée, Marion Duval de Yvan Pommaux.

Le texte dans les bulles est court (68 mots) et occupe essentiellement les 6 premières cases. Le vocabulaire utilisé ne présente pas de difficulté particulière, il relève d'un style direct et courant. Le support de la bande dessinée est facilitateur car l'écrit est étayé par l'image.

Question

Marion pourra se prélasser devant la télé avec un « super sandwich » ce soir parce que...

FIGURE 31 – Exemple question groupe <1 et 1

| | | | |
|---|-------------------------------------|--|--------|
| 1 | <input type="checkbox"/> | <i>Son papa va lire le journal.</i> | 3,3 % |
| 2 | <input checked="" type="checkbox"/> | <i>Son papa travaille.</i> | 89,7 % |
| 3 | <input type="checkbox"/> | <i>Son papa lui a préparé un sandwich.</i> | 1,8 % |
| 4 | <input type="checkbox"/> | <i>On ne peut pas savoir.</i> | 5 % |

L'élève en réussite a utilisé les deux informations présentes dans les bulles des deux premières cases de la planche :

- le papa sera absent le soir car il a à faire au journal ;
- Marion doit se débrouiller toute seule.

La proximité des éléments favorise l'établissement d'un lien entre ceux-ci malgré l'absence de connecteur logique dans le texte (du type « parce que » ou « car »). La réussite à cet item est élevée : près de 9 élèves sur 10 (89,7 %) accèdent à la bonne réponse.

5.3.2 Item caractéristique du groupe 2

Le groupe 2 représente 26,6 % des élèves évalués. Il est constitué des élèves capables de prélever des informations et de faire des inférences dans des écrits plus longs où ils sont capables d'identifier une chaîne référentielle.

Support

FIGURE 32 – Exemple support groupe 2

Lorsqu'elle allait au marché, ma mère me laissait au passage dans la salle de classe de mon père, qui apprenait à lire à des gamins de six ou sept ans. Je restais assis, bien sage, au premier rang et j'admirais la toute-puissance paternelle. Il tenait à la main une baguette de bambou: elle lui servait à montrer les lettres et les mots qu'il écrivait au tableau noir.

Un beau matin, ma mère me déposa à ma place, et sortit sans mot dire, pendant qu'il écrivait magnifiquement sur le tableau: «La maman a puni son petit garçon qui n'était pas sage.»

Tandis qu'il arrondissait un admirable point final, je criai:«Non! Ce n'est pas vrai!»

Mon père se retourna soudain, me regarda stupéfait, et s'écria:«Qu'est-ce que tu dis?

- Maman ne m'a pas puni! Tu n'as pas bien écrit!... »

Texte littéraire de 151 mots extrait de La gloire de mon père de Marcel Pagnol. Le contexte et le lexique sont courants puisque la scène se déroule dans une salle de classe.

Question

FIGURE 33 – Exemple question groupe 2

Lis cet extrait du texte.

« Il tenait à la main une baguette de bambou... »

Dans cet extrait du texte, le mot écrit en gras remplace...

- | | | | |
|---|-------------------------------------|----------------------|--------|
| 1 | <input type="checkbox"/> | <i>le narrateur.</i> | 5,3 % |
| 2 | <input type="checkbox"/> | <i>les gamins.</i> | 4,1 % |
| 3 | <input type="checkbox"/> | <i>ma mère.</i> | 1,9 % |
| 4 | <input checked="" type="checkbox"/> | <i>mon père.</i> | 87,9 % |

La compétence visée porte sur l'identification d'une reprise anaphorique proche : il s'agit d'associer le pronom personnel « il » à son personnage. Le référent « père » est présent deux lignes au-dessus. Il y a également une reprise du mot « père » à travers l'utilisation de l'adjectif qualificatif « paternelle » qui se trouve juste avant le pronom personnel.

La plupart des élèves répondent de façon correcte à cette question avec un taux de réussite de presque 88 %.

5.3.3 Item caractéristique du groupe 3

Le groupe 3 représente près d'un tiers (32,1 %) des élèves évalués. Il est constitué des élèves capables d'accéder à la compréhension globale d'un écrit long et d'affiner leur appréhension des personnages et du narrateur. Ils construisent des inférences à partir d'informations dispersées dans un texte narratif.

Support

Texte littéraire long (voir Annexe) extrait du roman *Le Hollandais sans peine* de Marie Aude Murail, présent dans la liste de référence des ouvrages littéraires du cycle 3, niveau de lecture 1 (sur 3). Il évoque un contexte accessible aux élèves, celui des vacances, les personnages principaux ont l'âge des élèves. Néanmoins, un des personnages principaux est aussi le narrateur, ce qui représente une difficulté dans le système énonciatif. L'humour est très présent dans ce texte. Cela engage le lecteur dans le traitement implicite de certaines informations à prendre au second degré. Enfin si le lexique et la syntaxe relèvent d'un niveau courant, la présence de mots inventés en complexifie cependant la compréhension.

Question

Coche les autres façons d'appeler Jean-Charles et Niclausse dans le texte.

FIGURE 34 – Exemple question groupe 3

| | Jean-Charles | | Niclausse | |
|--------------------------|-------------------------------------|-------|-------------------------------------|-------|
| Moatazan | <input checked="" type="checkbox"/> | 64,9% | <input type="checkbox"/> | 23% |
| Le petit garçon blond | <input type="checkbox"/> | 16% | <input checked="" type="checkbox"/> | 73,3% |
| Le petit garçon allemand | <input type="checkbox"/> | 12,9% | <input checked="" type="checkbox"/> | 76% |
| Le narrateur | <input checked="" type="checkbox"/> | 72% | <input type="checkbox"/> | 10,9% |
| Mon nouveau camarade | <input type="checkbox"/> | 14,2% | <input checked="" type="checkbox"/> | 71,6% |
| je | <input checked="" type="checkbox"/> | 72,8% | <input type="checkbox"/> | 11,2% |
| Moi, Tarzan | <input checked="" type="checkbox"/> | 73% | <input type="checkbox"/> | 15,4% |
| Un excellent professeur | <input checked="" type="checkbox"/> | 62,2% | <input type="checkbox"/> | 24,5% |

Les réponses proposées reprennent des mots du texte sous forme de groupes nominaux éparpillés dans l'ensemble du texte. Une des réponses correspond à la proposition « le narrateur » (qui n'apparaît donc pas en tant que mot dans le texte) et enfin une autre correspond au pronom personnel « je » qui se retrouve tout au long du texte. Pour réussir cet item, l'élève doit se construire une représentation mentale correcte du texte, intégrer les caractéristiques des personnages et s'appuyer sur le contexte pour établir le lien entre les personnages et leur reprise anaphorique. La gymnastique de va-et-vient dans ce texte plutôt long est conséquente et coûteuse sur le plan cognitif. Deux sous-items sont moins réussis que les autres, car relevant de compétences plus fines : l'identification de Jean-Charles comme « Moatazan » et « un excellent professeur ». En effet « Moatazan » est un mot inventé. Sa construction est assez rocambolesque puisqu'il s'agit du prénom que s'adjuge Jean-Charles pour le reste des vacances auprès de son camarade. A un moment Jean-Charles se présente en se montrant et en disant « Moi Tarzan » mais Niclausse le déforme en répétant « Moatazan ». Il est donc ici question d'un détournement d'identité qui peut échapper à la compréhension de certains lecteurs. Quant à « un excellent professeur », il s'agit d'un sous-item dont la position en fin d'exercice et en fin de texte renforce la difficulté. En outre il faut reconnaître, derrière les termes « un excellent

professeur », une synthèse réalisée de façon humoristique par Jean-Charles lui-même sur ses propres compétences pédagogiques. Là encore une compréhension fine du texte est nécessaire et marque ainsi le niveau de compétences des élèves appartenant au groupe 3.

Le seuil de réussite de ce tableau série est fixé à 7 sous-items réussis sur les 8. Tout comme pour le groupe 2, l'exercice porte sur les reprises anaphoriques. Néanmoins le niveau de difficulté est plus élevé comme l'indique le taux de réussite qui est de 60,7%.

5.3.4 Item caractéristique du groupe 4

Le groupe 4 est constitué de plus d'un cinquième des élèves évalués (21,2 %). Il s'agit des élèves présentant une très bonne maîtrise des connaissances et compétences exigibles en fin d'école. Ils sont capables de développer des stratégies expertes et de mobiliser leur attention sur un texte long pour en atteindre une représentation globale et accéder aux intentions de l'auteur.

Support

Le support est un texte littéraire extrait du roman *Poil de Carotte* de Jules Renard (voir Annexe). Sa difficulté réside moins dans la longueur (507 mots) que dans son contenu. Le champ sémantique est celui de la vie quotidienne à la campagne de la fin du XIXe siècle. Le contenu du texte est donc plutôt éloigné du quotidien des élèves. Le vocabulaire employé est soutenu, il comporte des termes qui n'ont plus cours : « couardise, poltron, indolent » mais qui font l'objet de renvois en notes de bas de page avec des explications. La structure du texte est complexe. La mise en situation est abrupte, le texte débute directement par des dialogues et met en avant des personnages qui ne sont pas présentés. On ne comprend les relations entre les personnages qu'au fur et à mesure.

Question

Félix et Ernestine ne veulent pas sortir pour aller fermer les poules dans le poulailler parce qu'il fait jour. Ils envoient donc Poil de carotte le faire car lui, il n'a pas peur du noir. Le jeune garçon sort courageusement et réussit à enfermer les poules. À son retour, il reçoit les félicitations de toute la famille.
Demain ce sera le tour de Félix de le faire

Dans ce texte, des mots ou groupes de mots sont soulignés. Indique dans le

tableau ci-dessous s'ils sont en logique avec le texte de départ ou non.

FIGURE 35 – Exemple question groupe 4

| | oui | | non | |
|---|-------------------------------------|-------|-------------------------------------|-------|
| Félix et Ernestine ne veulent pas sortir | <input checked="" type="checkbox"/> | 84,1% | <input type="checkbox"/> | 6,8% |
| parce qu'il fait jour | <input type="checkbox"/> | 12,4% | <input checked="" type="checkbox"/> | 77,8% |
| il n'a pas peur du noir | <input type="checkbox"/> | 30,9% | <input checked="" type="checkbox"/> | 58,7% |
| il réussit à enfermer les poules | <input checked="" type="checkbox"/> | 83,4% | <input type="checkbox"/> | 6,9% |
| il reçoit les félicitations de toute la famille | <input type="checkbox"/> | 29,3% | <input checked="" type="checkbox"/> | 60,9% |
| Demain, ce sera le tour de Félix | <input type="checkbox"/> | 21,8% | <input checked="" type="checkbox"/> | 68,4% |

L'exercice proposé part d'une réécriture résumée du texte d'origine. Certaines propositions soulignées dans ce résumé sont vraies, d'autres sont fausses. Pour réussir, l'élève doit développer des compétences synthétiques de compréhension de l'organisation du texte. Il doit se représenter l'histoire, retenir les informations principales et enfin comparer avec ce qui est proposé pour repérer la cohérence ou l'incohérence des propositions avec le texte d'origine. La capacité de mémorisation de l'élève est importante pour être efficace dans la réalisation de l'exercice. Le seuil de réussite est fixé à 6/6. L'exercice présenté ici n'est réussi que par un peu plus de la moitié des élèves (51,2%).

5.3.5 Item caractéristique du groupe 5

Le groupe 5 comprend 9,2 % des élèves évalués. Ces élèves présentent une maîtrise parfaite des connaissances et compétences exigibles en fin d'école. Leur réussite passe notamment par la compréhension de l'évolution des sentiments des personnages.

Support

Les élèves découvrent un texte narratif complet et long (774 mots) : La petite boule d'argile de Diana Engel, comportant trois illustrations. L'histoire évoque un groupe d'enfants participant à une activité. Le contexte peut ne pas être connu (il s'agit du travail de l'argile en poterie), il ne fait cependant pas appel à

un champ lexical spécialisé et la langue est d'un niveau standard. En revanche, on a ici un texte où des élèves peuvent rencontrer des difficultés au niveau du système énonciatif, puisqu'il est question d'une petite boule d'argile personnifiée et douée de sentiments. De plus, c'est ensuite le point de vue d'une petite fille qui va utiliser cette boule qui est exprimé.

Question

Décris les différents sentiments ressentis par la boule d'argile au début et à la fin de l'histoire. Explique pourquoi ses sentiments ont changé.

FIGURE 36 – Exemple question groupe 5

| | |
|-----------|--|
| Exemple 1 | <p>Au début la boule d'argile est triste car personne ne l'a manipulée.</p> <p>À la fin, elle est joyeuse est fière car elle sent enfin à quelque chose.</p> |
| Exemple 2 | <p>Au début elle était triste, à la fin elle est contente.</p> |
| Exemple 3 | <p>Au début la boule a été choisie par la fille et à la fin elle a été transformée en tasse.</p> |
| Exemple 4 | <p>Elle est terrifiée car elle est restée sur le bord de la fenêtre elle puis elle a durci mais la pluie est arrivée donc elle est redevenue molle.</p> |

Il s'agit d'une question ouverte avec correction experte.

L'élève doit ici effectuer de multiples inférences de sentiments portant sur l'ensemble du texte. Les capacités en jeu, à savoir identifier et se remémorer l'évolu-

tion des sentiments ainsi que déterminer la cause de cette évolution nécessitent d'avoir une vision synthétique sur un texte long, marquant la parfaite maîtrise des compétences de lecture de ces élèves du groupe 5. Le taux de réussite est plutôt faible : moins d'un tiers des élèves ont réussi cet item. Cela s'explique aussi certainement par le double questionnement et le fait qu'il fallait enchaîner deux réponses correctes pour considérer l'item réussi. La production 1 présentée ici montre un exemple de bonne réponse avec les deux éléments bien pris en compte : les sentiments et les explications. Les exemples 2, 3 et 4 montrent des éléments erronés. Dans l'exemple 2, l'élève n'a pris en compte que la première partie de la consigne en ne décrivant que les sentiments de la boule d'argile. Dans l'exemple 3, l'élève décrit des éléments de narration de façon erronée (la boule n'est choisie par la fille qu'à la fin), cela montre que la compréhension n'est que parcellaire. Enfin, dans l'exemple 4, le début de la production est correct (avec une prise en compte du sentiment de la petite boule d'argile et de sa raison). En revanche, la fin de la réponse montre que l'élève s'est perdu en cours de route. Il s'est arrêté sur la transformation de l'état physique de la boule. Il parle du fait que la boule est redevenue molle sans décrire de sentiments. Surtout, l'élève ne va pas jusqu'au bout de l'histoire avec le fait que la boule est enfin façonnée en tasse et qu'elle en éprouve une très grande fierté.

La difficulté de cette question aboutit à un taux de réussite assez faible avec moins d'un tiers des élèves (29,6 %).

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 du secteur d'enseignement. Le lecteur est invité à consulter la Note d'Information pour plus de détails (Eteve et al., 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 écoles, 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 les cycles antérieurs. Pour chaque établissement des échantillons de 2003, 2009, 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 18).

TABLEAU 18 – Score moyen selon l'indice de position sociale moyen de l'établissement (CEDRE maîtrise de la langue école 2003, 2009, 2015 et 2021)

| Indice moyen école | Année | Répartition (%) | Score moyen | Écart type |
|--------------------|-------|-----------------|-------------|------------|
| 1er quart | 2003 | 24,8 | 231 | 49 |
| 1er quart | 2009 | 23,8 | 238 | 48 |
| 1er quart | 2015 | 24,6 | 235 | 42 |
| 1er quart | 2021 | 24,8 | 240 | 43 |
| 2e quart | 2003 | 25,1 | 247 | 49 |
| 2e quart | 2009 | 25,6 | 249 | 46 |
| 2e quart | 2015 | 25,2 | 250 | 43 |
| 2e quart | 2021 | 24,4 | 252 | 45 |
| 3e quart | 2003 | 24,9 | 256 | 48 |
| 3e quart | 2009 | 24,4 | 259 | 48 |
| 3e quart | 2015 | 24,9 | 252 | 41 |
| 3e quart | 2021 | 25,8 | 254 | 44 |
| 4e quart | 2003 | 25,1 | 266 | 47 |
| 4e quart | 2009 | 26,2 | 258 | 49 |
| 4e quart | 2015 | 25,3 | 265 | 41 |
| 4e quart | 2021 | 25 | 271 | 46 |

Note de lecture : en 2021, le score moyen des élèves appartenant au quart des écoles accueillant les élèves les plus favorisées (quatrième quart) est de 271 point et l'écart-type est de 46 points. Les évolutions significatives sont indiquées en gras.

6.2 Élaboration des questionnaires de contexte

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

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

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

test.

Le questionnaire élève contient aussi un certain nombre de questions à renseigner par l'enseignant(e), il s'agit des questions concernant la catégorie socioprofessionnelle des parents mais aussi le parcours de l'élève (raccourcissement de cycle ou maintien dans un cycle, orientation retenue etc.).

6.3 Finalités assignées à la lecture par les élèves en 2021

Adossés aux exercices d'évaluation, des questionnaires de contexte étaient aussi proposés aux élèves, les interrogeant notamment sur leurs pratiques de lecture. En dehors de l'école, 58,9% élèves déclarent lire pour apprendre alors qu'ils sont 45,8% à lire pour s'évader et 44,1% pour s'informer (figure 37). Le rapport à la lecture varie selon l'origine sociale des élèves. Ainsi, 68,1% des élèves appartenant au quart des écoles les moins favorisées déclarent lire pour apprendre. Parmi ceux appartenant aux écoles les plus favorisées, ils sont 52,3% dans ce cas. Ces derniers sont, en revanche, beaucoup plus nombreux à déclarer lire pour s'évader (56,4% contre 39,1% dans les écoles les moins favorisées)(figure 38). C'est aussi plus souvent le cas chez les filles que chez les garçons (52,9% contre 43,8%)(figure 39).

FIGURE 37 – Finalités assignées à la lecture par les élèves en 2021 (en %)

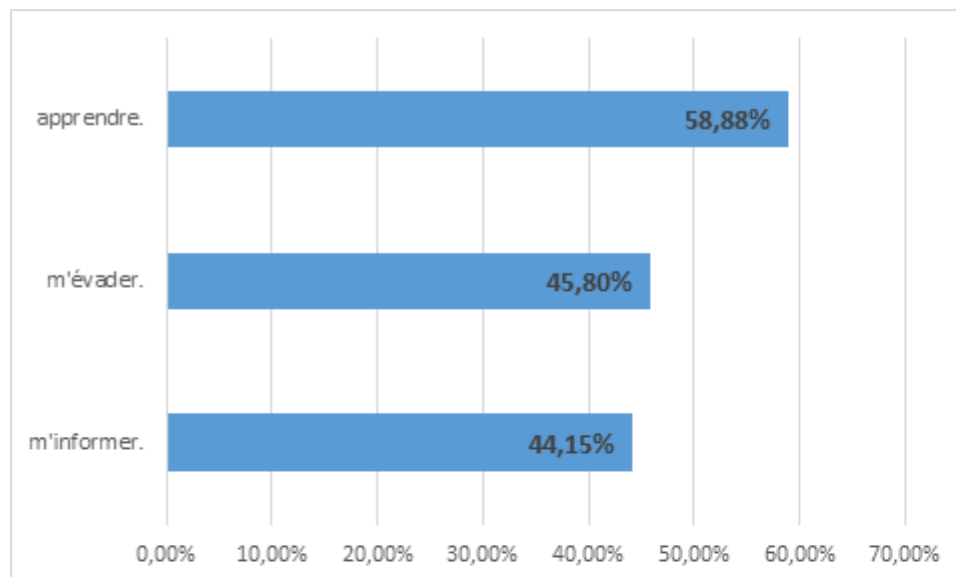


FIGURE 38 – Finalités assignées à la lecture par les élèves en 2021 par indice de position sociale (en %)

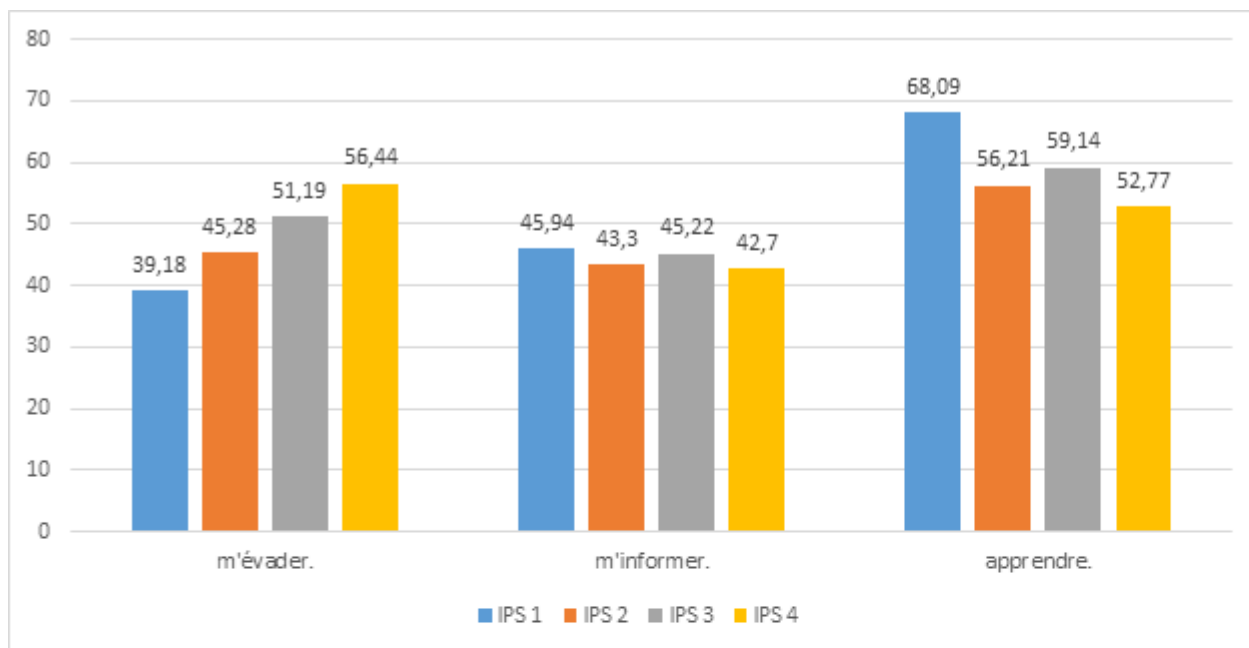
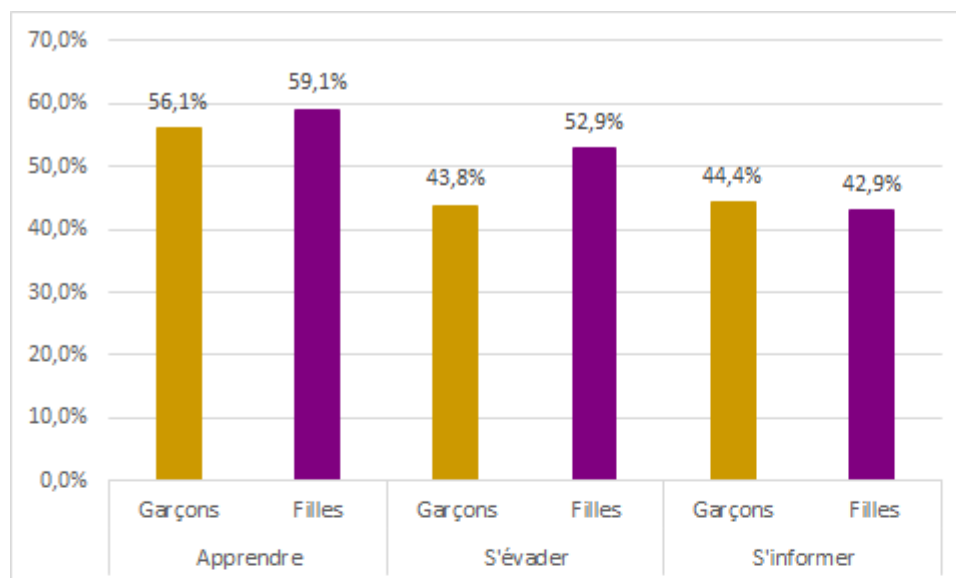


FIGURE 39 – Finalités assignées à la lecture par les élèves en 2021 par sexe (en %)



7 Annexe

Certification AFNOR pour les évaluations CEDRE

La DEPP est engagée dans un processus de certification. Les évaluations CEDRE sont certifiées depuis mars 2015.

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.

7.1 Textes utilisés dans l'évaluation

7.1.1 Le Hollandais sans peine

Au dîner, mon père m'expliqua comment je devais prendre mon bain de langue :

- Tu vas faire connaissance avec un petit garçon allemand de ton âge. Vous jouerez ensemble, il te dira des mots en allemand, tu les répéteras et ça viendra tout seul.

Je bougonnai :

- Je n'ai pas envie de jouer avec un garçon allemand.

Maman s'écria :

- Les enfants allemands sont aussi bien que les enfants français !

- Non, ils sont bêtes, dis-je.

Mon père prit de nouveau son air des grands jours :

- Jean-Charles, tu me fais de la peine. Les enfants ont tous la même valeur, qu'ils soient blancs ou noirs, espagnols ou allemands.

Je répétais tout bas :

- Ils sont bêtes.

Mais vraiment tout bas, pour ne pas déclencher une tempête.

C'est alors qu'une dame blonde accompagnée d'un petit garçon blond passèrent devant notre tente. Ils portaient la vaisselle du dîner dans deux cuvettes. La dame nous regarda, sourit et nous dit quelque chose.

- Bonsoir ! claironnèrent Papa et Maman.

Le petit garçon nous jeta un coup d'œil. Il avait mon âge, il était probablement allemand, il campait à deux pas de nous.

- Et tu vois, dit mon père, il aide sa maman à faire la vaisselle.

- Propose-lui une partie de ballon, ajouta ma mère.

Mes parents me regardaient, ma sœur me regardait, les voisins de tente me regardaient, même le chien du gardien me regardait. La terre entière attendait que j'aie joué au ballon avec le petit garçon allemand. Je haussai les épaules, je donnai un coup de pied dans mon ballon et je me dirigeai en ronchonnant vers la tente d'à côté. Le petit garçon semblait m'attendre, les mains sur les hanches. Je shootai. Il arrêta mon ballon sans effort. Il était sûrement idiot, mais il n'était pas maladroit. La partie s'engagea.

Au bout de dix minutes, j'avais oublié mon bain de langue, mais je m'amusais bien. Le petit garçon blond bloqua le ballon sous son pied et, tapant sur sa poitrine, il me cria :

- Niclausse !

Où quelque chose dans ce genre-là. Je compris qu'il se présentait. Je tapai sur ma poitrine et, pour plaisanter, je criai :

- Moi, Tarzan !

Mon nouveau camarade était un enfant sérieux. Il répéta après moi :

- Moatazan.

Il avait l'intention de prendre un bain de langue. Il répéta une deuxième fois : « Moatazan », avec beaucoup d'application. Je n'aimais pas tellement mon prénom. Je songeai que « Moatazan » ferait tout aussi bien l'affaire que « Jean-Charles » pendant ce mois d'août.

Nous nous assîmes dans l'herbe. L'idée me traversa alors qu'il est très difficile de devenir copain de quelqu'un qui ne dit pas un seul mot de la même façon que vous. Mon ami Niclausse cueillit une fleur et prononça un mot comme « flour » ou « flaour » ou peut-être « flaveur ». Par politesse, je répétais. Il éclata de rire. J'avais sans doute mal prononcé. Il me fit signe ensuite de donner son nom à la fleur dans ma langue.

Que se passa-t-il dans ma tête, à ce moment-là ? Je trouvai tout à coup stupide d'appeler « fleur » une fleur. Je savais bien que c'était son nom ! Alors, je dis :

- Chprout !

Niclausse répéta :

- Chprout.

C'était sûrement un bon élève à l'école. Je secouai la tête pour indiquer qu'il prononçait mal. Je rectifiai :

- Chprouout !

Et Niclausse répéta. Pris soudain d'une véritable folie, je lui montrai un arbre :

- Trabeun !

- Trabeun, dit Niclausse.

Puis, pour ne pas perdre le vocabulaire fraîchement acquis, il récapitula :

- Chprouout, trabeun !

Je fis bravo en tapant dans mes mains. Puis je lui désignai notre toile de tente :

- Chrapati.

- Chrapati, fit Niclausse, élève docile.

Au bout du dixième mot, je pris peur de tout mélanger, d'autant que Niclausse était doué d'une mémoire redoutable. Je courus à ma tente en criant :

- Moatazan chrapati...

Ce qui signifiait, bien évidemment, que j'allais faire un saut jusqu'à ma tente.

Niclausse me comprit parfaitement. Mes parents me virent entrer, tout excité.

- Tu t'amuses bien ? me demanda Papa.

- Oh oui ! Je vais écrire des mots dans un cahier.

Je me saisis du fameux cahier de vacances.

- Des mots d'allemand ? me demanda mon père, plein d'espoir.

- Non, criai-je en m'enfuyant, c'est du hollandais ! Niclausse est hollandais !

J'étais ravi de cette dernière trouvaille.

Le lendemain après-midi, j'avais déjà recouvert six feuilles de mon cahier de vacances avec un lexique franco-hollandais. Comme j'étais un excellent professeur, Niclausse progressait très vite. En fin de journée, nous avons presque des petites conversations. Je disais :

- Moatazan gaboum chrouillasse.

Ce qui voulait dire :

- Moatazan aimer mer.

Niclausse me répondait avec conviction :

- Niclausse gaboum chrouillasse.

Mon ami était persuadé qu'il apprenait le français et je crois bien, que de temps en temps, il allait répéter un mot de français à ses parents.

7.1.2 Poil de Carotte

- Je parie, dit madame Lepic, qu'Honorine a encore oublié de fermer les poules. C'est vrai. On peut s'en assurer par la fenêtre. Là-bas, tout au fond de la grande cour, le petit toit aux poules découpe, dans la nuit, le carré noir de sa porte ouverte.
- Félix, si tu allais les fermer ? dit madame Lepic à l'aîné de ses trois enfants.
- Je ne suis pas ici pour m'occuper des poules, dit Félix, garçon pâle, indolent¹ et poltron².
- Et toi, Ernestine ?
- Oh ! Moi, maman, j'aurais trop peur !
- Grand frère Félix et sœur Ernestine lèvent à peine la tête pour répondre. Ils lisent, très intéressés, les coudes sur la table, presque front contre front.
- Dieu, que je suis bête ! dit madame Lepic. Je n'y pensais plus. Poil de Carotte, va fermer les poules ! Elle donne ce petit nom d'amour à son dernier né, parce qu'il a les cheveux roux et la peau tachée. Poil de Carotte, qui joue à rien sous la table, se dresse et dit avec timidité :
- Mais, maman, j'ai peur aussi, moi.
- Comment ? répond madame Lepic, un grand gars comme toi ! C'est pour rire. Dépêchez-vous, s'il te plaît !
- On le connaît ; il est hardi comme un bouc, dit sa sœur Ernestine.
- Il ne craint rien ni personne, dit Félix, son grand frère.
- Ces compliments enorgueillissent Poil de Carotte, et, honteux d'en être indigne, il lutte déjà contre sa couardise³. Pour l'encourager définitivement, sa mère lui promet une giflette.
- Au moins, éclairez-moi, dit-il.
- Madame Lepic hausse les épaules, Félix sourit avec mépris.
- Seule pitoyable⁴, Ernestine prend une bougie et accompagne petit frère jusqu'au bout du corridor.
- Je t'attendrai là, dit-elle.
- Mais elle s'enfuit tout de suite, terrifiée, parce qu'un fort coup de vent fait vaciller la lumière et l'éteint.
- Poil de Carotte, les fesses collées, les talons plantés, se met à trembler dans les ténèbres. Elles sont si épaisses qu'il se croit aveugle. Parfois une rafale l'enveloppe, comme un drap glacé, pour l'emporter. Des renards, des loups même, ne lui soufflent-ils pas dans ses doigts, sur sa joue ? Le mieux est de se précipiter, au juger, vers les poules, la tête en avant, afin de trouver l'ombre. Tâtonnant, il saisit le crochet de la porte. Au bruit de ses pas, les poules effarées s'agitent en gloussant sur leur perchoir. Poil de Carotte leur crie : « Taisez-vous donc, c'est moi ! » ferme la porte et se sauve, les jambes, les bras comme ailés. Quand il rentre, haletant, fier de lui, dans la chaleur et la lumière, il lui semble qu'il échange des loques⁵ pesantes de boue et de pluie contre un vêtement neuf et léger. Il sourit, se tient droit, dans son orgueil, attend les félicitations, et maintenant hors de danger, cherche sur le visage de ses parents la trace des inquiétudes qu'ils ont eues. Mais grand frère Félix et sœur Ernestine continuent tranquillement leur lecture, et madame Lepic lui dit, de sa voix naturelle :
- Poil de Carotte, tu iras les fermer tous les soirs.

d'après Jules Renard, Poil de Carotte

¹ mou, sans énergie

² peureux, lâche

³ peur

⁴ qui éprouve de la pitié

⁵ vieux habits

La petite boule d'argile

Diana Engel

Au sommet d'une vieille tour, il y avait un atelier. C'était un atelier de poterie, rempli de pots d'émail de toutes les couleurs, de tours de potier, de fours et, bien entendu, d'argile. Près de la fenêtre, il y avait un grand coffre de bois avec un lourd couvercle. C'est là qu'on conservait l'argile. Tout au fond, aplatie dans un coin, se trouvait la plus vieille boule d'argile. Il y avait bien longtemps qu'on l'avait manipulée, elle s'en rappelait à peine. Chaque jour, le lourd couvercle était soulevé. Des mains venaient prendre des sacs ou des morceaux d'argile. La petite boule entendait les sons joyeux des potiers occupés à leur travail.

« Quand donc viendra mon tour ? », se demandait-elle. À mesure que les jours passaient, la petite boule d'argile perdait espoir dans l'obscurité du coffre.

Un jour, enfin, un grand groupe d'enfants vient à l'atelier avec un maître. Plusieurs petites mains plongent dans le coffre. La petite boule d'argile est choisie en dernier, mais au moins, elle est sortie.

« Voilà ma chance ! », pensa-t-elle, clignant des yeux à cause de la lumière.

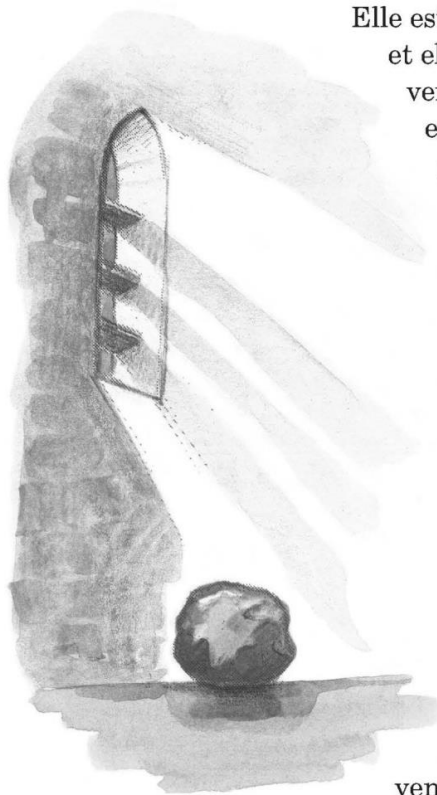
Un petit garçon la dépose sur le tour du potier qu'il fait tourner très, très vite. « C'est amusant ! », pensa la petite boule d'argile. Le petit garçon essaie d'étirer l'argile vers le haut, tandis que le tour tourne. Quel plaisir pour la petite boule d'argile de ressembler enfin à quelque chose ! Après avoir essayé de fabriquer un bol, le petit garçon abandonne. Puis il travaille l'argile pour en faire une boule bien ronde.

« C'est l'heure de ranger », dit le maître. L'atelier se remplit du bruit que font les enfants qui épongent, raclent, lavent et essuient. L'eau goutte partout.

Le petit garçon dépose la boule d'argile près de la fenêtre et court rejoindre ses amis. Au bout d'un moment, l'atelier est vide. La pièce est silencieuse et sombre. La petite boule d'argile est effrayée. Non seulement l'humidité du coffre lui manque, mais en plus elle sait qu'elle est en danger.

« C'est la fin », pense-t-elle. « Je vais rester ici, sécher et devenir dure comme une pierre. »





Elle est là, près de la fenêtre ouverte, incapable de bouger et elle sent l'humidité la quitter. Le soleil tape, puis le vent de la nuit souffle et la petite boule d'argile devient dure comme une pierre. Elle est si dure qu'elle peut à peine réfléchir. Tout ce qu'elle sait c'est qu'il n'y a plus d'espoir.

Mais quelque part au fond d'elle, il reste une petite goutte d'humidité, qu'elle refuse de laisser aller.

« De la pluie », pense-t-elle.

« De l'eau », soupire-t-elle.

« Par pitié », laisse-t-elle finalement échapper de son corps sec et désespéré.

Un nuage qui passe par là, a pitié de la petite boule d'argile et une chose merveilleuse se produit alors. De grosses gouttes de pluie, passant par la fenêtre ouverte, tombent sur elle avec force. Il pleut toute la nuit et au matin elle est redevenue aussi molle qu'avant.

Des voix se font entendre dans l'atelier.

« Oh non ! », dit une potière qui travaille souvent à l'atelier. « Quelqu'un a laissé la fenêtre ouverte !

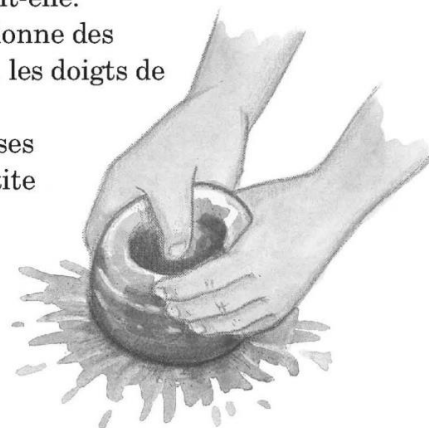
Il faut nettoyer tout ça. Tu peux prendre un peu d'argile le temps que j'aille chercher des torchons. », dit-elle à sa fille.

La petite fille voit la petite boule d'argile sur le bord de la fenêtre.

« Voilà un morceau parfait pour moi », se dit-elle.

Bientôt, elle presse et pétrit l'argile et lui donne des formes amusantes. Pour la petite boule d'argile, les doigts de la petite fille sont magiques.

La petite fille réfléchit tout en travaillant, ses mains s'activent dans un but bien précis. La petite boule d'argile se sent modelée avec douceur en une forme ronde et creuse. Quelques pressions et elle a une poignée.



« Maman, maman », crie la petite fille, « J'ai fait une tasse! »

« C'est magnifique! », dit sa mère. « Pose-la sur l'étagère et je la cuirai au four. Tu pourras la décorer de la couleur de ton choix. »

Bientôt, la petite tasse est prête à être emportée dans sa nouvelle maison. Maintenant la petite boule d'argile vit sur une étagère dans la cuisine, à côté d'autres tasses, de bols et de soucoupes. Toutes sont très différentes et certaines sont très belles.

« À table! », dit la maman en posant la nouvelle tasse qu'elle remplit de chocolat chaud.

La petite fille la prend avec précaution. Comme la petite boule d'argile est fière des courbes délicates de sa nouvelle forme. Comme elle fait bien son travail!

La petite tasse se tient fièrement. « Enfin, je suis devenue quelque chose. »



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.
- Eteve, Y., Nghiem, X., & Philbert, L. (2021). CEDRE 2003 - 2009 - 2015 - 2021 - maîtrise de la langue en fin d'école : une légère hausse du niveau global des élèves en 2021. *Note d'information*, 22.
- Rocher, T. (1999). *Psychométrie et théorie des sondages* (Mémoire de Master non publié). Université Paris VI.
- Rocher, T. (2013). *Mesure des compétences : les méthodes se valent-elles ? questions de psychométrie dans le cadre de l'évaluation de la compréhension de l'écrit* (Thèse de doctorat non publiée). Université Paris-Ouest.
- Rocher, T. (2015). Mesure des compétences : méthodes psychométriques utilisées dans le cadre des évaluations des élèves. *Éducation et Formations*, 86-87, 37-60.
- Rocher, T. (2016). Construction d'un indice de position sociale des élèves. *Éducation et Formations*, 90, 5-27.
- Sautory, O. (1993). La macro calmar. redressement d'un échantillon par calage sur marges. *Série des documents de travail de l'INSEE, Document F9310*.
- Smith, R., Schumaker, R., & Bush, J. (1998). Using item mean squares to evaluate fit to the rasch model. *Journal of Outcome Measurement*, 2 n°1, 66-78.
- Tillé, Y. (2001). *Théorie des sondages. échantillonnage et estimation en populations finies. cours et exercices avec solution*. Paris : Dunod.
- Trosseille, B., & Rocher, T. (2015). Les évaluations standardisées des élèves. perspective historique. *Éducation et Formations*, 85-86, 15-35.
- Warm, T. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54 n°3, 427-450.

Liste des tableaux

| | | |
|----|--|----|
| 1 | Cycle des évaluations disciplinaires réalisées sur échantillons (CEDRE) depuis 2003 | 4 |
| 2 | Programme du cycle 3 - enseignement du Français (BOEN n°31 du 30 juillet 2020) | 5 |
| 3 | Compétences évaluées en compréhension de l'écrit dans l'évaluation Maîtrise de la langue 2021 (papier et numérique) | 6 |
| 4 | Compétences évaluées en étude de la langue dans l'évaluation Maîtrise de la langue 2021 (papier et numérique) | 7 |
| 5 | Compétences évaluées en compréhension de l'oral dans l'évaluation Maîtrise de la langue 2021 (numérique) | 7 |
| 6 | Déroulé de l'évaluation Maîtrise de la langue 2021 | 8 |
| 7 | Exclusions pour la base de sondage - CEDRE Maîtrise de la langue CM2 2021 | 36 |
| 8 | Répartition base de sondage - CEDRE Maîtrise de la langue CM2 2021 | 36 |
| 9 | Répartition dans l'échantillon - CEDRE Maîtrise de la langue CM2 2021 | 37 |
| 10 | Non-réponse des écoles - CEDRE Maîtrise de la langue CM2 2021 | 38 |
| 11 | Non-réponse globale - CEDRE Maîtrise de la langue CM2 2021 . | 38 |
| 12 | Comparaison entre les marges de l'échantillon et les marges dans la population | 40 |
| 13 | Scores moyens et erreurs standard associées - CEDRE Maîtrise de la langue CM2 | 40 |
| 14 | Répartition en % dans les groupes de niveaux - CEDRE Maîtrise de la langue CM2 | 41 |
| 15 | Erreurs standards des répartitions en % dans les groupes de niveaux - CEDRE Maîtrise de la langue CM2 | 41 |
| 16 | Effet du plan de sondage - CEDRE Maîtrise de la langue CM2 2015 | 42 |
| 17 | Niveaux de compétences CEDRE Maîtrise de la langue École (moyennes et écarts-type) | 60 |
| 18 | Score moyen selon l'indice de position sociale moyen de l'établissement (CEDRE maîtrise de la langue école 2003, 2009, 2015 et 2021) | 75 |

Table des figures

| | | |
|---|--|----|
| 1 | Répartition des types de textes selon le support proposé | 10 |
| 2 | Répartition des items de compréhension de l'écrit par compétences | 11 |
| 3 | Répartition des types de textes selon les compétences évaluées (en nombre d'items) | 12 |
| 4 | Répartition des items d'étude de la langue par compétences . . . | 13 |

| | | |
|----|---|----|
| 5 | Répartition des items d'étude de la langue par compétences | 14 |
| 6 | Exemple de format QCM | 15 |
| 7 | Exemple de format tableau série | 15 |
| 8 | Exemple de format champ libre | 16 |
| 9 | Répartition des items numériques par domaine | 17 |
| 10 | Répartition des types de textes selon le support proposé | 18 |
| 11 | Répartition des items numériques de compréhension de l'écrit par compétences | 19 |
| 12 | Répartition des items numériques de d'étude de la langue par compétences | 20 |
| 13 | Exemple d'item numérique de type phrase cylindre | 21 |
| 14 | Exemple d'item numérique de type menu déroulant | 22 |
| 15 | Exemple d'item numérique de type QCM | 22 |
| 16 | Exemple d'item numérique de type surligner | 23 |
| 17 | Exemple d'item numérique de type glisser déposer | 23 |
| 18 | Exemple d'item numérique de type pointer | 24 |
| 19 | Exemple d'item numérique de type champ libre | 24 |
| 20 | Répartition des items selon leur format | 25 |
| 21 | Exemple d'un support de lecture (bande dessinée) réparti sur 3 écrans. | 26 |
| 22 | Exemple d'un support de lecture (texte uniquement) réparti sur 2 écrans. | 28 |
| 23 | Exemple d'un item numérique de compréhension de l'oral | 29 |
| 24 | Taux de réussite des items en fonction de la position de leurs blocs | 43 |
| 25 | Représentation graphique utilisée pour le regroupement d'items . | 49 |
| 26 | Modèle de réponse à l'item - 2 paramètres | 52 |
| 27 | Exemples d'ajustements (FIT) | 56 |
| 28 | Comparaison des paramètres de difficulté 2015-2021 (CEDRE Maîtrise de la langue CM2) | 59 |
| 29 | Principes de construction de l'échelle | 62 |
| 30 | Exemple support groupe <1 et 1 | 65 |
| 31 | Exemple question groupe <1 et 1 | 66 |
| 32 | Exemple support groupe 2 | 67 |
| 33 | Exemple question groupe 2 | 67 |
| 34 | Exemple question groupe 3 | 69 |
| 35 | Exemple question groupe 4 | 71 |
| 36 | Exemple question groupe 5 | 72 |
| 37 | Finalités assignées à la lecture par les élèves en 2021 (en %) . . . | 77 |
| 38 | Finalités assignées à la lecture par les élèves en 2021 par indice de position sociale (en %) | 78 |

| | | |
|----|---|----|
| 39 | Finalités assignées à la lecture par les élèves en 2021 par sexe (en %) | 79 |
|----|---|----|