

# Documentation concernant les tests cognitifs réalisés par les enfants à domicile dans le cadre de l'enquête 10,5 ans dans Elfe et Epipage 2

Auteurs : Mélissa Arneton, Thierry Siméon, Laëtitia Marchand

Version du 26/03/2026

## Table des matières

1. Présentation générale des tests .....	2
2. Les variables mises à disposition .....	3
Calcul des scores .....	4
Matrices .....	4
Puzzles .....	5
PPVT .....	5
Mr Petit .....	6
3. Qualités psychométriques des différents tests cognitifs .....	8
Section A : Scores aux tests cognitifs dans Elfe .....	9
A.1 Distribution des scores .....	9
A.2 Analyses de fidélité .....	10
B.1 Distribution des scores .....	12
B.2 Analyses de fidélité .....	13
Références .....	14
Complément 1 - Un problème généralisé de chronométrage d'ordre aléatoire .....	16
Complément 2 – Mesure automatisée de la fluence par Agir pour l'école .....	17
Complément 3 - Éléments relatifs à la conversion des notes brutes en notes standardisées .....	20

## 1. Présentation générale des tests

Pour une présentation des informations recueillies auprès de l'enfant durant l'enquête à domicile à 10,5 ans voire la documentation disponible sur <https://plateforme-acces-donnees-elfe-france.site.ined.fr/fr/enquetes/8-10-ans>.

En 2022, dans le cadre de l'enquête à 10,5 ans, le protocole à domicile incluait des épreuves cognitives à réaliser par l'enfant sur une tablette. Présentées comme une succession de petits jeux de logique sur une application dédiée, quatre épreuves cognitives communément utilisées dans les travaux portant sur le développement de l'enfant étaient proposées à 7597 enfants Elfe et 1490 enfants Epipage 2 : une épreuve de compréhension d'une histoire, présentée à l'enfant comme « Monsieur Petit », une épreuve de complément de matrices « Dessins », une épreuve de rotation mentale « Puzzles » et une épreuve de vocabulaire « Trouve les mots ».

Une application numérique dédiée a été créée et paramétrée à partir de différents subtests cognitifs, dans le respect des droits d'auteurs et de propriété intellectuelle. La solution a été conçue pour fonctionner sans connexion à internet pendant le recueil et pour être compatibles avec une tablette tournant sur Windows 10. Au total, 9087 enfants ont été interrogés à domicile par 196 enquêteurs. Le fait d'utiliser une tablette numérique avait pour but d'homogénéiser les recueils, de faciliter le lancement du protocole par les enquêteurs et de présenter un outil appétant pour les enfants afin de maintenir leur attention. Le rôle de l'enquêteur s'est limité à lancer successivement les différents tests et à veiller à ce que l'enfant ne soit pas dérangé pendant le test. Une fois le logiciel lancé, l'enfant était autonome dans sa progression, jusqu'à l'arrêt programmé du test en fonction du taux de réussite de l'enfant. Les consignes étaient affichées à l'écran et lues à l'enfant par un enregistrement audio. Le pilotage du développement informatique associé a été assuré pour Elfe par Franck Ramus et Ghislaine Labouret du laboratoire de Sciences Cognitives et Psycholinguistique (CNRS / ENS).

Les quatre tâches cognitives sont des épreuves complexes, nécessitant de l'enfant d'inférer des règles de fonctionnement puis de les appliquer pour résoudre les items. Concernant la compréhension d'un texte à lire, il s'agissait du texte « Monsieur Petit » de la batterie EL.F.E élaborée au sein du groupe académique de pratique et de recherche Cognisciences<sup>1</sup>, ce subtest est nommé Mr Petit dans la présente note. Pour les subtests Matrices et Puzzles, les items sont repris de la version Q-interactive du WISC-5<sup>2</sup> et ce avec l'autorisation de la maison d'édition Pearson. Pour le test de vocabulaire, les items sont repris de la version numérique du *Peabody Picture Vocabulary Test*<sup>3</sup> (PPVT-5) distribué par la maison d'édition Pearson. Les stimuli des tests issus du WISC-V et du PPVT-5 sont la propriété exclusive de l'éditeur Pearson, qui a accordé une autorisation de les utiliser uniquement dans le cadre de cette étude.

Le format numérique permet d'adapter les items présentés à l'enfant en fonction des items auxquels il a répondu correctement précédemment et ce, en conformité avec les pratiques adoptées dans les versions numériques des tests présentés aux enfants (Braga Filaho et al., 2021 ; Daniel et al., 2014 ; Gilbert et al., 2021). Les tests sont réalisés dans un délai indicatif ou en durée limitée pour d'autres. Le temps global prévu pour l'ensemble des tests est d'une trentaine de minutes et au maximum d'1 heure 30 (temps moyen observé de 29 minutes dans Elfe). En termes d'adaptation aux besoins particuliers, la présentation des items en fonction des réponses précédentes de l'enfant contribue à rendre accessible l'outil aux enfants présentant des troubles cognitifs. L'ergonomie choisie avec une adaptation des

---

<sup>1</sup> Batterie d'évaluation de la lecture en fluence de 2008 mis à disposition ([Cognisciences](#)).

<sup>2</sup> WISC-V - Échelle d'intelligence de Wechsler pour enfants et adolescents - 5ème édition ([pearsonclinical.fr](#)).

<sup>3</sup> PPVT-5 – Évaluation du vocabulaire réceptif ([pearsonclinical.fr](#)).

images à la taille de l'écran, une taille de police lisible, un affichage épuré ainsi qu'à une présentation des consignes de chaque test à l'écrit et en version audio contribue à une accessibilité de l'outil. Pour les enfants ayant des difficultés dans l'usage autonome de l'outil (en raison d'une déficience visuelle ou motrice par exemple), l'enquêteur a pu apporter une aide en reportant leurs réponses sur l'ordinateur. De plus certains tests prévoient le recueil d'information complémentaire, ainsi dans le PPVT, une variable spécifique du test indique si l'enfant se déclare comme sourd ou mal-entendant. Cinq enfants seulement l'ont mentionné. En raison du caractère identifiant d'une telle variable, elle ne sera pas mise à disposition en tant que telle, il conviendra de demander plutôt une variable plus générale sur les troubles de l'audition renseignée par ailleurs (intitulé HISTRAUD10).

Un certain nombre de dysfonctionnements ont été observés durant la passation par les enquêteurs en raison de bugs de l'outil (voir Complément 1). L'analyse des données montre que le chronométrage n'a pas fonctionné pour un certain nombre d'enfants, avec des dépassements du temps alloué pour répondre : pour le test des matrices pour 5 % des enfants ELFE et 2,7% des enfants EPIPAGE ; pour les puzzles visuels pour 31 % des enfants ELFE et 13.1% des enfants EPIPAGE. Ces dépassements non conformes aux règles d'application des tests ont été corrigés (voir complément 1).

La présente note rend compte des variables disponibles suite aux arbitrages effectués pour chacun des tests Matrices, Puzzles, PPVT ainsi que pour l'épreuve de fluence. Concernant la lecture du texte « Monsieur Petit », un codage des erreurs de lecture a été réalisé par intelligence artificielle pour 7124 enfants Elfe et 1421 enfants Epipage 2. Un codage manuel complémentaire a été réalisé pour les enfants pour lesquels l'IA ne parvenait pas à rendre compte des résultats en raison d'un débit trop rapide ou de bruits parasites (118 enfants Elfe et 62 enfants Epipage 2). L'analyse des résultats et la description des scores sont incluses dans la note.

## 2. Les variables mises à disposition

Les bonnes réponses à chaque test sont établies conformément aux protocoles de cotation des outils utilisés ; une réponse donnée après le délai imparti dans les règles d'application est assimilée à une réponse fautive. Pour les matrices, les puzzles et le PPVT, trois types de variables sont fournies : le score obtenu fondé sur le calcul des réponses correctes, la note standardisée sur l'âge selon les étalonnages des différents outils (voir complément 2). Pour Mr Petit, six types de variables sont fournies : la note standardisée est sur la classe de scolarisation attendue à 10 ans soit le CM2 conformément à l'étalonnage EL.F.E., le score brut, le nombre total d'erreur ainsi que le type d'erreurs. **Seules les variables synthétiques sont mises à disposition de la communauté des chercheurs** ; l'accès aux résultats des enfants item par item peut faire l'objet d'une demande auprès de l'équipe projet : [secretariat-cad-elfe@ined.fr](mailto:secretariat-cad-elfe@ined.fr).

Les variables sont indexées comme suit dans la base de données :

- ⇒ A10COG\_WISC5MATSCORE\_STD : Note standardisé au test des matrices
- ⇒ A10COG\_WISC5MATSCORE\_C : Score brut au test des matrices
- ⇒ A10COG\_WISC5MATSTATUS : Statut du test des matrices en trois modalités (fait de manière complète, incomplète ou non fait)
- ⇒ A10COG\_WISC5PUZSCORE\_STD : Note standardisé au test des puzzles
- ⇒ A10COG\_WISC5PUZSCORE\_C : Score brut au test des puzzles
- ⇒ A10COG\_WISC5PUZSTATUS : Statut du test des puzzles en trois modalités (fait de manière complète, incomplète ou non fait)
- ⇒ A10COG\_PPVTSCORE\_STD : Note standardisé au test au PPVT
- ⇒ A10COG\_PPVTSCORE\_C : Score brut au PPVT

- ⇒ A10COG\_PPVTSTATUS : Statut du test PPVT en trois modalités (fait de manière complète, incomplète ou non fait)
- ⇒ A10F\_MPETIT\_outlier : Qualité du codage en trois modalités (qualité suffisante pour un codage par IA, qualité suffisante pour un codage manuel, non fait).
- ⇒ A10F\_MPETIT\_MCLM\_CALC : Estimation du nombre de mots correctement lus par minute.

Pour le test M. Petit codé par IA, quatre variables supplémentaires peuvent être fournies.

- ⇒ A10F\_MPETIT\_N\_ERROR : Nombre d'erreurs total
- ⇒ A10F\_MPETIT\_DELETION : Nombre d'erreurs de type suppression
- ⇒ A10F\_MPETIT\_SUBSTITUTION : Nombre d'erreurs de type substitution
- ⇒ A10F\_MPETIT\_INSERTION : Nombre d'erreurs de type insertion

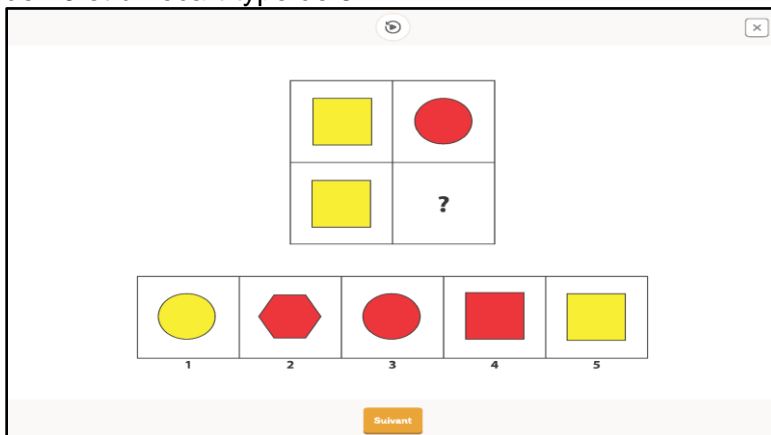
Dans le catalogue de données Elfe, une variable supplémentaire est disponible : A10F\_MPETIT\_MCLM\_STD catégorise le score de l'enfant selon la classe de scolarisation de référence en CM2 (normal, inférieur, difficultés importantes en lecture) cf complément 3.

### Calcul des scores

Le score d'un test est calculé pour les enfants dont on considère qu'ils ont bien participé au test en question ; on retire ainsi, pour les calculs, les enfants pour lesquels l'enquêteur ou l'enquêtrice a signalé un bug empêchant la participation de l'enfant ainsi que les réponses enregistrées qui ne respectent pas les règles d'arrêt établis dans la littérature. Seuls les scores sur ces variables corrigées sont présentés ici.

### Matrices

Le test des matrices mesure l'aptitude de l'enfant au raisonnement logique visuospatial, c'est-à-dire à détecter la relation conceptuelle sous-jacente entre des objets visuels et à utiliser le raisonnement afin d'identifier et d'appliquer les règles. Le test se compose d'une série de figures abstraites (27 séries dont 2 séries d'essai soit 34 items). Il s'agit pour l'enfant de retrouver, parmi un éventail de figures, celle qui complète la série. Parmi 5 figures proposées, l'enfant clique sur celle qui lui semble correspondre le mieux, il a 60 secondes maximum pour répondre par item. La première matrice posée à tous est la numéro 5. En cas d'erreur sur la matrice 5 ou 6, la matrice 4 est présentée, si la réponse est erronée la matrice 3 est présentée et ainsi de suite. Dans le cadre de deux réponses correctes consécutives, le protocole présente une matrice de niveau plus élevée : par exemple après une réussite aux matrices 5 et 6, la matrice 7 est présentée. Les enfants ayant réussi la matrice 5 et suivantes se voit attribué les points correspondant à la réussite des matrices 1, 2, 3 et 4. Le nombre de séries présentées est déterminé par le nombre de bonnes réponses de l'enfant. Après trois réponses fausses consécutives, le test s'arrête. Le score correspond au nombre d'items correctement répondus, il est au maximum de 32. La note standardisée varie de 1 à 19 ; avec une moyenne de 10 et un écart-type de 3.



### Références pour aller plus loin :

Gonthier, C. (2022). Cross-cultural differences in visuo-spatial processing and the culture-fairness of visuo-spatial intelligence tests: an integrative review and a model for matrices tasks. *Cogn. Research*, 7 (11). <https://doi.org/10.1186/s41235-021-00350-w>

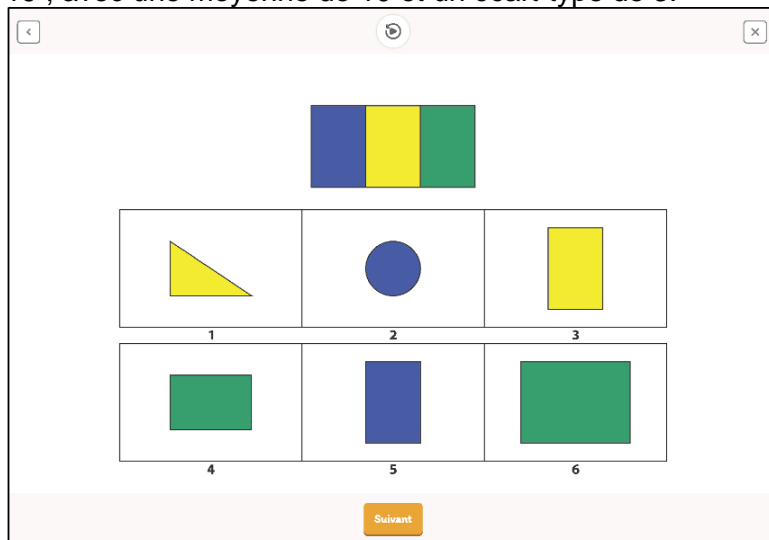
Laporte, E. et Terriot, K. (2018). WISC-V : une évolution attendue et nécessaire. *L'orientation scolaire et professionnelle*, 47(1) [en ligne]. <https://doi.org/10.4000/osp.5658>

Raiford, S., Drozdick, L. et Zhang, O. (2015). *Q-interactive © Special Group Studies: The WISC © –V and Children with Autism Spectrum Disorder and Accompanying Language Impairment or Attention-Deficit/Hyperactivity Disorder*. Report, 11, Pearson Edition.

Wechsler, D. (2016). *WISC-V – échelle d'intelligence de Werchsler pour enfants et adolescents – 5<sup>ème</sup> édition*. ECPA par Pearson.

### Puzzles

Le test des puzzles renseigne sur les compétences visuospatiales, c'est-à-dire sur la capacité de l'enfant à analyser et synthétiser des éléments visuels pour former un tout cohérent. L'enfant doit choisir des images qui vont ensemble, proposées initialement sur 2 lignes distinctes (association de 2 images) puis sur 3 lignes distinctes (association de 3 images). Après deux items d'entraînement, 29 items sont à réaliser chacun dans un délai maximum de 30 secondes. Les items sont présentés par niveau de difficulté ascendante. La passation à 10 ans débute par le puzzle numéro 5. Si l'enfant répond correctement, il passe à l'item suivant, et les items 1 à 4 sont comptabilisés comme corrects. Si l'item 5 est échoué, les items précédents sont présentés par niveau de difficulté décroissante. La présentation des items s'arrête après 3 erreurs consécutives. Le score est le nombre de réponses correctes sur les items 1 à 29, le score aux Puzzles est au maximum de 32. La note standardisée varie de 1 à 19 ; avec une moyenne de 10 et un écart-type de 3.



### Références pour aller plus loin :

Laporte, E. et Terriot, K. (2018). WISC-V : une évolution attendue et nécessaire. *L'orientation scolaire et professionnelle*, 47(1) [en ligne]. <https://doi.org/10.4000/osp.5658>

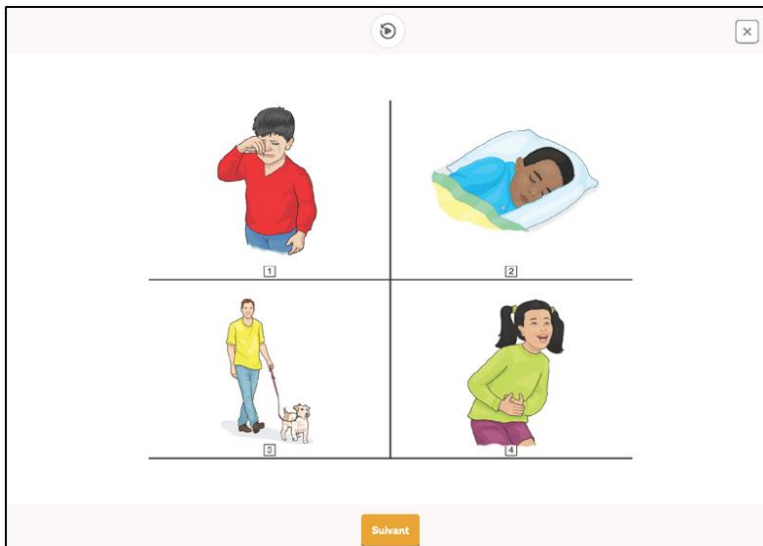
Raiford, S., Drozdick, L. et Zhang, O. (2015). *Q-interactive © Special Group Studies: The WISC © –V and Children with Autism Spectrum Disorder and Accompanying Language Impairment or Attention-Deficit/Hyperactivity Disorder*. Report, 11, Pearson Edition.

Wechsler, D. (2016). *WISC-V – échelle d'intelligence de Werchsler pour enfants et adolescents – 5<sup>ème</sup> édition*. ECPA par Pearson.

### PPVT

Le test est conçu pour mesurer le vocabulaire compris en fonction de l'âge, il s'agit à la fois noms d'objets, d'actions ou de concepts plus abstraits. L'enfant entend un mot préenregistré et doit choisir parmi 4 images celle qui correspond le mieux. Après 5 items d'entraînement, le test se poursuit avec 203 items ordonnés par niveau de complexité. Le temps alloué pour

répondre à chaque item est de maximum 30 secondes. A l'âge de 10 ans, le test débute à l'item 79. Le test s'arrête lorsque l'enfant a échoué 6 fois à 8 items consécutifs. Si l'enfant ne résout pas l'item 79, l'item 78 lui est présenté : s'il réussit il reprend à l'item 80, s'il échoue les items précédents lui sont proposés par ordre décroissant de difficulté. Le score total correspond aux nombres de réponses correctes sur l'ensemble des 203 items qu'ils soient ou non réalisés par l'enfant. Pour les enfants ayant réussi l'item 79, sont considérés comme acquis les items 1 à 78, sauf si erreurs initiales. Le score au PPVT est au maximum de 203. L'étalonnage du PVVT est établi sur des enfants entendants locuteurs du français en langue 1. La note standardisée varie de 40 à 160, avec une moyenne de 100 et un écart-type de 15.



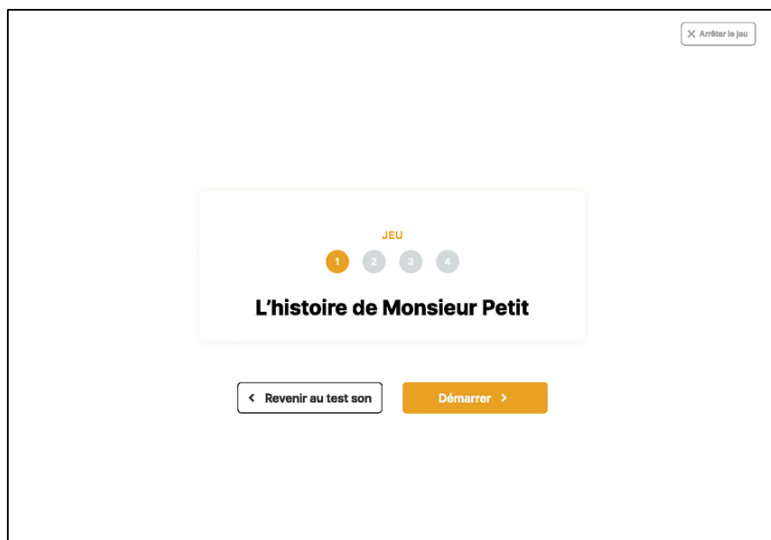
Références pour aller plus loin :

Armstrong, R. et al. (2016). Predicting receptive vocabulary change from childhood to adulthood: A birth cohort study. *Journal of Communication Disorders*, 64, 78-90.  
<https://doi.org/10.1016/j.jcomdis.2016.10.002>

Dunn, D. M (2019). PPVT5. *Échelle de vocabulaire en images Peabody – cinquième édition*. Manuel. Pearson édition.

### Mr Petit

Il s'agit d'un test de fluence de la lecture, c'est-à-dire de « la capacité à lire avec aisance, rapidement, sans erreur et avec une intonation adaptée » dans un temps limité. L'enfant lit à voix haute le maximum d'un texte en une minute. Sa lecture est enregistrée afin d'établir ensuite un score en nombre de mots correctement lus par minute. Si l'enfant hésite ou se reprend pour finir par lire correctement un mot, il ne s'agit pas d'une erreur. Seuls les mots ou phrases non lus ou sautés sont comptabilisés en erreur. Les hésitations conduisant à une bonne lecture limitent le nombre de mots lus dans la durée d'une minute. Trois types d'erreur sont identifiées : l'oubli d'un mot (suppression), le remplacement d'un mot par un autre (substitution) et l'ajout d'un mot supplémentaire non présent dans le texte (insertion). Le test de Mr Petit contient 365 mots. Le score maximal au test de fluence est de 365. L'étalonnage de l'outil par classe de scolarisation indique qu'en classe de CM2 le score moyen obtenu est de 137 mots lus avec un écart-type de 31 ; 10% des enfants lisent 177 mots ou plus en 1 minute. Les scores mis à disposition dans Elfe et Epipage 2 sont issus d'un partenariat relatif au développement d'un logiciel d'automatisation du codage mené par l'équipe de Franck Ramus et de l'association Agir pour l'Ecole et du codage manuel complémentaire réalisé par l'équipe de Franck Ramus.



Références pour aller plus loin :

Juton, A. et Lequette, C. (n.d.). *Évaluation de la fluence en lecture, du CP au lycée. A partir des outils cognoscience : OURA, ELFE, ECLA 16+*. [En ligne] <https://www1.ac-grenoble.fr/media/14939/download>

Lequette, C., Pouget, G., & Zorman, M. (2014). *Fluence - volume 3, Les petits guides CM*. Grenoble : Les Éditions de la Cigale.

Zorman, M., Lequette, C., Pouget, C., Devaux, M. F., & Savin, H. (2008). Entraînement de la fluence de lecture pour les élèves de 6<sup>e</sup> en difficulté de lecture. *A.N.A.E*, 96-97, 213-220.

**Selon les enfants, entre 1 et 4 scores sont disponibles, étant donné que certains enfants n'ont pas réalisé les quatre tests** (voir tableau suivant). La notion de taux de couverture (tableau ci-dessous) indique la proportion d'enfants ayant participé parmi tous ceux susceptibles de répondre à 10 ans.

Tableau 1 : Effectifs et taux de couverture selon le test cognitif et l'échantillon<sup>4</sup>

	Effectifs ( $n_{Elfe}$ ; $n_{Epi\text{page}}$ )	Taux de couverture ( $p_{Elfe}$ ; $p_{Epi\text{page}}$ )
<b>Matrices</b>	7321 ; 1426	.9975 ; .957
<b>Puzzles</b>	7319 ; 1425	.9975 ; .956
<b>PPVT</b>	7292 ; 1415	.9941 ; .950
<b>Mr Petit</b>	7242 ; 1483	.9845 ; .954

<sup>4</sup> Pour le test Mr Petit, sont considérés dans cette version de la documentation, les enfants dont les données ont pu être traitées par IA.

### 3. Qualités psychométriques des différents tests cognitifs

Conformément aux usages psychométriques et afin de pouvoir situer les performances obtenues par les enfants Elfe et Epipage 2 aux échantillonnages existant sur ces tests, les scores seront décrits en termes de caractéristiques de tendance centrale (moyenne et médiane) et de variations (écart-type et quartiles). Des graphes relatifs à l'allure générale des courbes des scores bruts avant standardisation illustrent les descriptions.

Pour les tests Matrices, Puzzles et PPVT réalisés en passation par ordinateur, la cohérence interne est présentée par deux indicateurs : l'alpha de Cronbach et l'omega de McDonald (calculés avant d'établir les notes standardisées). L'alpha de Cronbach indique dans quelle mesure les items constitutifs d'une échelle vont ensemble et relèvent d'une même dimension. Cependant, il est sensible aux nombres d'items et aux échantillons de grande taille. Le coefficient de fidélité omega de McDonald est indicatif de la force de l'association entre les items et le score représentatif du construit théorique mesuré. Les indices varient de 0 à 1, une valeur supérieure à 0.7 sera considérée comme un indice d'ajustement acceptable. Pour le test Mr Petit, qui n'est pas constitué d'items, le processus de codage par IA est décrit dans le complément 2 de la documentation. Pour 180 enfants dont les enregistrements n'étaient pas traitables par l'IA (118 enfants Elfe et 62 enfants Epipage 2), deux évaluatrices ont réalisé un codage en double aveugle<sup>5</sup>.

En plus des indices de validité interne à chaque test, deux analyses relatives à la validité externe des tests en tenant compte de la littérature sont présentées : une analyse des corrélations inter-tests et une analyse selon une variable traditionnellement investiguée dans la littérature, le genre. Les tests cognitifs mesurent à la fois des dimensions spécifiques (fluence, compréhension verbale, capacités visuospatiales, raisonnement logique) et un élément commun générale de capacités cognitives. Cette variable générale peut être identifiée par des corrélations entre les scores aux différents tests. Une description des scores selon le genre est indiquée pour compléter ce panorama global de la qualité des scores dans les échantillons.

Pour chacun des échantillon Elfe (section A) et Epipage 2 (section B), les indicateurs sont calculés lorsque le statut du test est considéré comme complet, c'est-à-dire que l'arrêt a eu lieu car l'enfant avait atteint son score maximum de compétences. Les graphiques et tableaux suivants présentent les distributions (non-commentées) des scores.

À compter de la vague de recueil à 10,5 ans, les données collectées par Elfe et Epipage 2 sont harmonisées dès la procédure de collecte, les chercheurs peuvent donc compiler en une base commune, les données demandées auprès du comité d'accès aux données d'Elfe et du comité d'accès aux données d'Epipage. Pour les données collectées aux vagues précédentes, les chercheur·e·s doivent identifier les variables qu'ils souhaitent harmoniser (un catalogue des variables harmonisées est en cours de constitution).

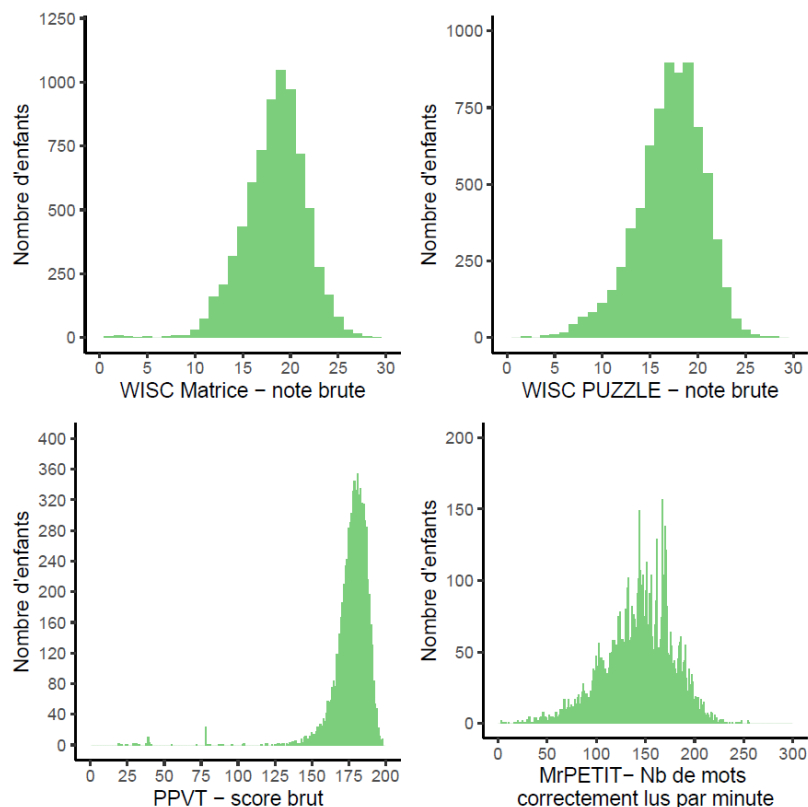
---

<sup>5</sup> Quand leur codage différait de moins de cinq points, la moyenne entre les deux dates est indiquée en base. Pour 18 enfants, une écoute commune a été réalisée pour établir le nombre de mots corrects lus par l'enfant et entendus.

## Section A : Scores aux tests cognitifs dans Elfe

**7328 enfants** (3679 filles et 3811 garçons) ont participé à au moins un des tests de manière complète, 7209 ont les quatre tests (Matrices, Puzzles, PPVT, et Mr Petit) complets.

Figure 1 : Allure des courbes sur les scores bruts observés



### A.1 Distribution des scores

Tableau 2 : Statistiques descriptives de l'ensemble de l'échantillon ELFE

Test	N Complet	Médiane [écart interquartile]	Moyenne (écart- type)
<b>Matrices</b>			
Note brute	7321	19 [16-20]	18.4 (3.2)
Note standard	<b>7321</b>	<b>11 [8-12]</b>	<b>10.2 (2.4)</b>
<b>Puzzles</b>			
Note brute	7319	17 [15-19]	17.1 (3.5)
Note standard	<b>7319</b>	<b>11 [9-12]</b>	<b>10.8 (2.4)</b>
<b>PPVT</b>			
Score brut	7292	179 [173-185]	176.8 (14.8)
Score standard	<b>7292</b>	<b>100 [93-108]</b>	<b>100.0 (12.8)</b>
<b>Mr Petit</b>			
Note brute	7242	146 (123-168)	143.4 (35.0)

Note de lecture : 7321 enfants ont fini le test des matrices ; après standardisation sur l'âge, 25% des enfants ont un score inférieur ou égal à 8 et 25% ont un score supérieur ou égal à 12. Le score moyen standardisé à ce test est de 10.2 avec un écart-type de 2.4.

L'étalonnage à Mr Petit indique que 80% des enfants Elfe ont des performances normales comparativement à des enfants scolarisés en CM2. Dans les commentaires, au PPVT, 3 enfants sont mentionnant comme malentendants et à Mr Petit, 5 enfants sont indiqués comme dyslexiques.

## A.2 Analyses de fidélité

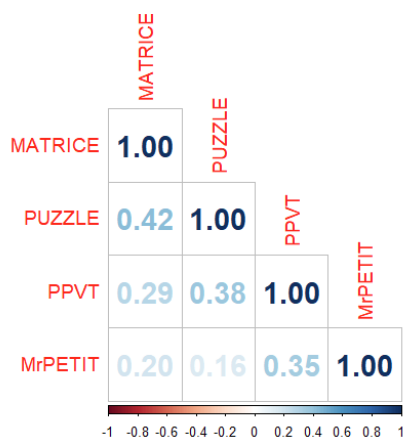
Figure 2 : Analyse de consistance interne

Test	Nombre d'items	Alpha de Cronbach	Omega	Taille de l'échantillon
<b>Matrices</b>	29	0.76	0.81	7321
<b>Puzzles</b>	32	0.81	0.83	7319
<b>PPVT</b>	203	0.94	0.99	7292

Les indices de consistance interne des items de chacun des trois tests, calculés à partir des données brutes quand le test est considéré comme complet, sont satisfaisants.

Pour le test Mr Petit, qui n'est pas constitué d'items, la qualité du codage par IA est décrite dans le complément 2, l'étalonnage repris de l'outil est présenté dans le complément 3.

Figure 3 : Corrélations entre les tests cognitifs

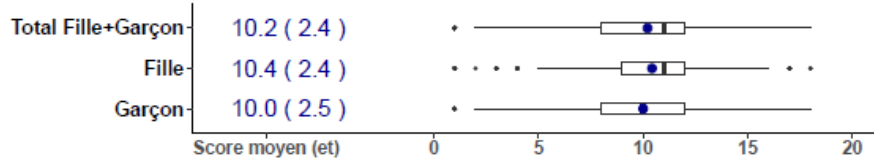


Les tests sont faiblement liés entre eux, les corrélations varient de .16 entre Mr Petit et les puzzles à .43 entre les puzzles et les matrices. Les tests mesurent des dimensions différentes du fonctionnement cognitif conformément à ce qui observé dans la littérature. Les corrélations sont plus importantes entre d'un côté les tests relevant d'une dimension visuospatiale (puzzles et matrices) et de l'autre les tests liés au langage (PPVT et test de fluence de lecture avec Mr Petit).

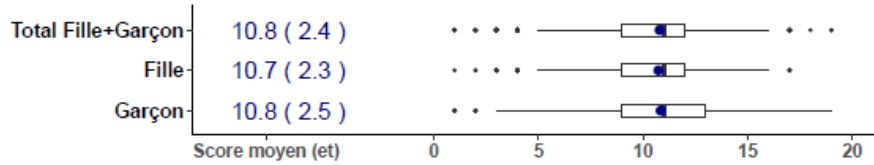
L'analyse descriptive selon le genre (figure 4 ci-dessous) n'indique pas un effet systématique à l'avantage des garçons ou des filles y compris pour des tests présentant un indice de corrélation supérieur à .30. Aux tests des matrices, le score moyen des filles est légèrement supérieur à celui des garçons (10,4 avec un écart-type de 2,4 vs 10 avec un écart-type de 2,5). Aux puzzles, les filles obtiennent un score moyen légèrement inférieur à celui des garçons (10,7 avec un écart-type de 2,3 vs 10,8 avec un écart-type de 2,5). Au PPVT, les filles obtiennent un score moyen un peu inférieur à celui des garçons (99,6 avec un écart-type de 12,4 vs 100,3 avec un écart-type de 13,1). Au test de fluence de lecture Mr Petit, le score brut moyen des filles est supérieur à celui des garçons (146,3 avec un écart-type de 34,6 vs 140,6 avec un écart-type de 35,1). En considérant le CM2 comme classe de scolarisation la plus fréquente à 10 ans, les filles et les garçons ont des compétences en lecture correspondant aux attendus à ce niveau scolaire (voir complément 3). De manière générale, les performances des garçons présentent une dispersion plus importante que celles des filles.

Figure 4 : Analyse descriptive selon le genre

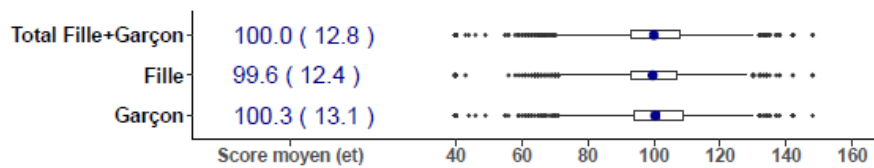
**MATRICE- Note standard**



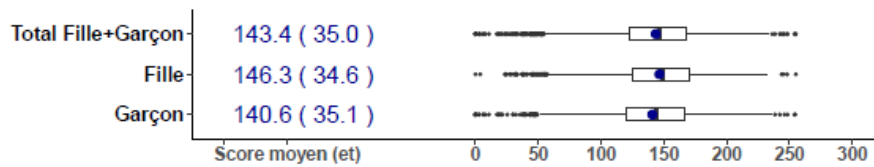
**PUZZLE- Note standard**



**PPVT- Score standard**



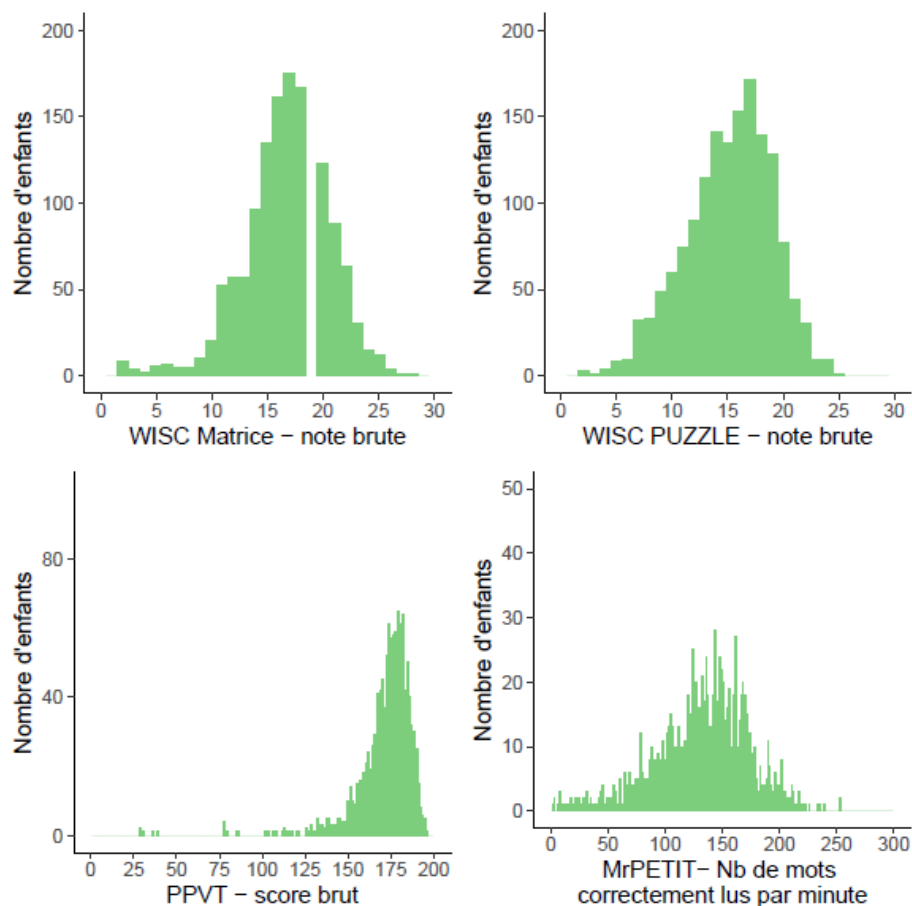
**MrPETIT- nb de mots lus**



## Section B : Scores aux tests cognitifs dans Epipage 2

1519 enfants (714 filles et 805 garçons) ont participé à au moins un des tests (Matrices, Puzzles, PPVT et Mr Petit) et 1476 de manière complète (Epipage + recrutement complémentaire).

Figure 5 : Allure des courbes sur les scores bruts observés



### B.1 Distribution des scores

Tableau 3 : Statistiques descriptives de l'ensemble de l'échantillon EPIPAGE (tronc commun)

Test	N Complet	Médiane [écart interquartile]	Moyenne (écart- type)
<b>MATRICE</b>			
Note brute	1426	17 [15-19]	16.9 (3.8)
Note standard	<b>1426</b>	<b>9 [7-11]</b>	<b>9.0 (2.8)</b>
<b>PUZZLE</b>			
Note brute	1425	16 [13-18]	15.1 (3.9)
Note standard	<b>1425</b>	<b>10 [8-11]</b>	<b>9.4 (2.6)</b>
<b>PPVT</b>			
Score brut	1415	175 [167-182]	171.8 (17.1)
Score standard	<b>1415</b>	<b>95 [86-103]</b>	<b>93.7 (14.2)</b>
<b>Mr Petit</b>			
Note brute	1483	136 [107 – 160]	132.0 (40.0)

Note de lecture : 1426 enfants ont fini le test des matrices ; après standardisation sur l'âge, 25% des enfants ont un score inférieur ou égal à 7 et 25% ont un score supérieur ou égal à 11. Le score moyen à ce test est de 9.0 avec un écart-type de 2.8.

L'étalonnage à Mr Petit indique que 60% des enfants Epipage ont des performances normales comparativement à des enfants scolarisés en CM2. Dans les commentaires, au PPVT, 1 enfant est indiqué comme malentendant.

## B.2 Analyses de fidélité

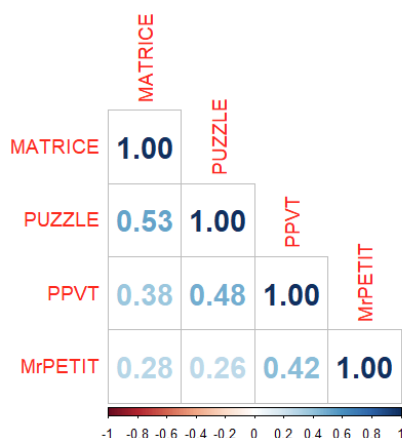
Figure 6 : Analyse de consistance interne

Test	Nombre d'items	Alpha de Cronbach	Omega	Taille de l'échantillon
<b>Matrices</b>	29	0.83	0.85	1426
<b>Puzzles</b>	32	0.84	0.86	1425
<b>PPVT</b>	203	0.95	0.98	1415

Les indices de consistance interne de chacun des trois tests, calculés à partir des données brutes quand le test est considéré comme complet, sont satisfaisants.

Pour le test Mr Petit, qui n'est pas constitué d'items, la qualité du codage par IA est décrite dans le complément 2 et l'étalonnage de référence dans le complément 3.

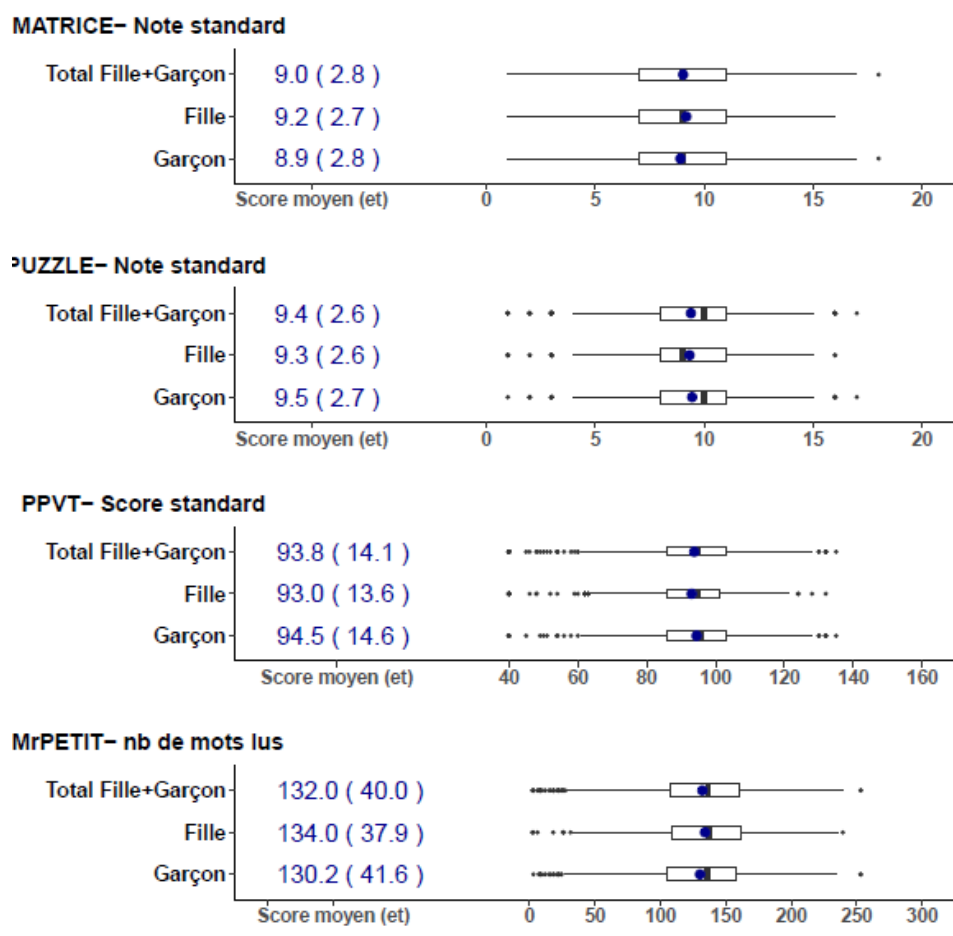
Figure 7 : Corrélations entre les tests cognitifs



Les tests sont liés entre eux, les corrélations varient de .26 entre Mr Petit et les puzzles à .53 entre les puzzles et les matrices. Le test de fluence de lecture est corrélé à .42 avec le test verbal du PPVT. Les puzzles sont un test de logique comprenant une dimension verbales, ils sont corrélés à .48 avec le PPVT et à .26 avec le test de fluence de lecture Mr Petit.

L'analyse descriptive selon le genre (figure 8 ci-dessous) n'indique pas un effet systématique à l'avantage des garçons ou des filles y compris pour des tests présentant un indice de corrélation supérieur à .30. Aux tests des matrices, le score moyen des filles est supérieur à celui des garçons (9,2 avec un écart-type de 2,7 vs 8,9 avec un écart-type de 2,8). Aux puzzles, les filles obtiennent un score moyen légèrement inférieur à celui des garçons (9,3 avec un écart-type de 2,6 vs 9,5 avec un écart-type de 2,7). Au PPVT, les filles obtiennent un score moyen inférieur à celui des garçons (93 avec un écart-type de 13,6 vs 94,5 avec un écart-type de 14,6). Au test de fluence de lecture Mr Petit, le score brut moyen des filles est supérieur à celui des garçons (134 avec un écart-type de 37,9 vs 130,2 avec un écart-type de 41,6). En moyenne, les filles et les garçons ont des compétences en lecture correspondant aux attendus en classe de CM2 (voir complément 3). De manière générale, les performances des garçons présentent une dispersion plus importante que celles des filles.

Figure 8 : Analyse descriptive selon le genre



## Références

- Armstrong, R. et al. (2016). Predicting receptive vocabulary change from childhood to adulthood: A birth cohort study. *Journal of Communication Disorders*, 64, 78-90. <https://doi.org/10.1016/j.jcomdis.2016.10.002>
- Braga Fialho, M., Rodrigues de Souza, E., & Mansur-Alves, M. (2021). Los efectos de las pruebas cognitivas computarizadas en el rendimiento de los niños con y sin síntomas de Trastorno de Déficit de Atención/Hiperactividad. *Ciencias Psicológicas*, 15(2), e-2393. <https://doi.org/10.22235/cp.v15i2.2393>
- Daniel, M. H., Whalstrom, D. & Zhang, O. (2014). *Equivalence of Q-interactive™ and Paper administration of Cognitive Tasks: WISC®-V*. Q-interactive technical report 8. <https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/wisc-v/q-interactive-wisc-v.pdf>
- Dunn, D. M (2019). PPVT5. *Échelle de vocabulaire en images Peabody – cinquième édition*. Manuel. Pearson édition.
- Gilbert, K., Kranzler, J. et Benson, N. (2021). An Independent Examination of the Equivalence of the Standard and Digital Administration Formats of the Wechsler Intelligence Scale for Children-5th Edition. *Journal of School Psychology*. 10.1016/j.jsp.2021.01.002
- Gonthier, C. (2022). Cross-cultural differences in visuo-spatial processing and the culture-fairness of visuo-spatial intelligence tests: an integrative review and a model for matrices tasks. *Cogn. Research*, 7 (11). <https://doi.org/10.1186/s41235-021-00350-w>

Hernández, A., Aguilar, C., Paradell, È., Muñoz, M. R., Vannier, L. C., & Vallar, F. (2017). The effect of demographic variables on the assessment of cognitive ability. *Psicothema*, 29(4), 469–474. <https://doi.org/10.7334/psicothema2017.33>

Juton, A. et Lequette, C. (2013.) *Evaluation de la fluence en lecture, du CP au lycée. A partir des outils cognoscences : OURA, ELFE, ECLA 16+*. [En ligne] <https://www1.ac-grenoble.fr/media/14939/download>

Laporte, E. et Terriot, K. (2018). WISC-V : une évolution attendue et nécessaire. *L'orientation scolaire et professionnelle*, 47(1) [en ligne]. <https://doi.org/10.4000/osp.5658>

Lequette, C., Pouget, G., & Zorman, M. (2014). *Fluence - volume 3, Les petits guides CM*. Grenoble : Les Éditions de la Cigale.

Raiford, S., Drozdick, L. et Zhang, O. (2015). *Q-interactive® Special Group Studies: The WISC®-V and Children with Autism Spectrum Disorder and Accompanying Language Impairment or Attention-Deficit/Hyperactivity Disorder*. Report, 11, Pearson Edition.

Wechsler, D. (2016). *WISC-V – échelle d'intelligence de Werchsler pour enfants et adolescents – 5<sup>ème</sup> édition*. ECPA par Pearson.

Zorman, M., Lequette, C., Pouget, C., Devaux, M. F., & Savin, H. (2008). Entraînement de la fluence de lecture pour les élèves de 6<sup>e</sup> en difficulté de lecture. *A.N.A.E*, 96-97, 213-220.

## Complément 1 - Un problème généralisé de chronométrage d'ordre aléatoire

Un certain nombre de dysfonctionnements ont été observés durant la passation, une étude a été menée pour identifier s'il s'agit d'un problème systématique de paramétrage des chronomètres, d'un problème de matériel lié à l'un des 196 enquêteurs ou d'un bug qu'il soit isolé ou systématique. L'analyse sur les temps moyens de réponse aux items indique que :

- aux matrices 6,39% des enfants ELFE et 2.7% des enfants EPIPAGE ont bénéficié de délai de réponse supérieur à 30 secondes pour certains items,
- aux puzzles 33,45% des enfants ELFE et 13.1% des enfants EPIPAGE ont bénéficié de délai de réponse supérieur à 30 secondes pour au-moins un item,
- le PPVT 0% des enfants ELFE et 1.1% des enfants EPIPAGE ont bénéficié de délai de réponse supérieur à 20 secondes pour certains items.

Le problème de chronométrage est généralisé, il ne semble pas dépendre du matériel utilisé par des enquêteurs spécifiques. Les temps de réponse à un item excédant le temps alloué sont supérieurs de plusieurs secondes aux durées prévues, il ne peut pas s'agir du temps de latence lié à l'outil informatique comme cela s'observe dans le chronométrage automatisé de l'enregistrement de la lecture de Mr Petit qui se fait à 1 minute et 1 seconde. La mise en relation des durées supérieures aux temps alloués par test et la mention de commentaires au moins une fois par les enquêteurs et enquêtrices indique qu'il n'y a pas de liens systématiques entre un matériel donné et le chronométrage défectueux. Les incidents de temps sont donc considérés comme un bug aléatoire. Ces résultats sont liés à des difficultés de programmation de l'outil numérique concernant les mesures des temps de présentation des items et des modalités temporelles de passation. Ces dépassements non conformes aux règles d'application des tests ont été corrigés. Afin de pouvoir tenir compte du délai de présentation de l'item et de la latence de quelques millisecondes constatées dans les applications, les temps de réponses sont majorés de 2 secondes. Ainsi sont considérés comme des bonnes réponses cotées 1, les réponses correctes intervenant dans le délai de 32\* secondes aux puzzles, de 32 secondes aux matrices et de 22 secondes au PPVT.

### Aperçu des distributions des scores suite aux recodages techniques des durées

Tableau 4 : Suivi des corrections dans ELFE

Test	Moyenne (écart-type) dans les données brutes	Moyenne (écart-type) score corrigé	Nombre d'enfants concernés
Matrices	18.43 (3.17)	18,36 (3.17)	289
Puzzles	17.29 (3.48)	17.05 (3.49)	261
PPVT	176,43 (17.03)	176.6 (15.11)	1

Tableau 5 : Suivi des corrections dans EPIPAGE

Test	Moyenne (écart-type) dans les données EPIPAGE	Moyenne (écart-type) score brut corrigé dans EPIPAGE	Nombre d'enfants concernés
Matrices	16.9 (3.8)	16.9 (3.8)	38
Puzzles	15.3 (3.9)	15.1 (3.9)	187
PPVT	171.5 (19.1)	171.8 (17.1)	15

## Complément 2 – Mesure automatisée de la fluence par Agir pour l'école

Le test Mr Petit a fait l'objet d'un partenariat relatif afin de mettre en œuvre une mesure automatisée par intelligence artificielle entre l'équipe de Franck Ramus pour Elfe et Epipage 2 et l'association Agir pour l'Ecole par Muy-Cheng Peich, Ludivine Reymbaut, Raphaël Deckers.

Pour toute demande d'information complémentaire, merci de contacter Agir pour l'école à l'adresse suivante : [mcpeich@agirpourlecole.org](mailto:mcpeich@agirpourlecole.org).

### 1. Description des données sources

Chaque enfant participant à l'étude lit le texte de référence « Monsieur Petit » sur un temps fixe d'une minute.

Un enregistrement au format \*.wav monocal 16khz a été réalisé lors de cette lecture. Pour analyser cet enregistrement et définir le nombre d'erreurs et leur type, un algorithme a été développé.

### 2. Traitement des données

L'algorithme a été développé pour le logiciel « Du Mot Au Texte » d'Agir pour l'école, puis adapté et spécialisé sur le contexte ELFE/EPIPAGE2 et le texte « Monsieur Petit ». 280 enregistrements ont été écoutés et codés manuellement et on servit à entraîner l'algorithme. Sur ces 280 échantillons, un test set de 80 échantillons a été conservé pour valider le comportement de l'automatisation sur les données en aveugle.

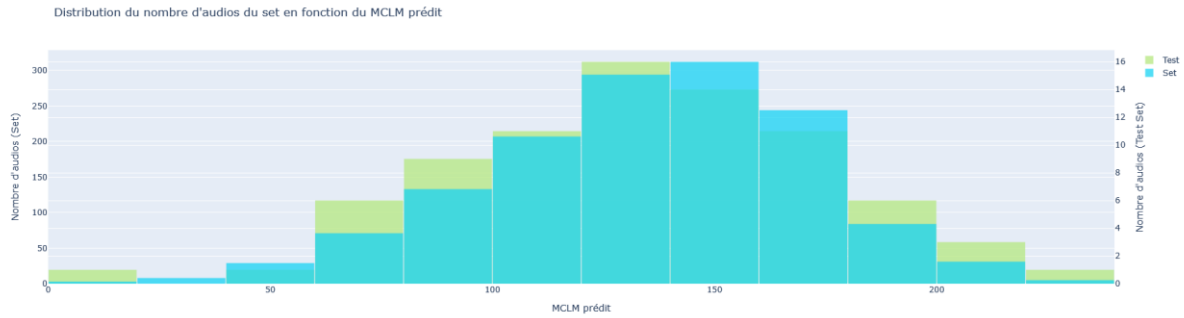
Le comportement a été validé sur la base de la distribution de la métrique principale (MCLM) entre le test set et les données en aveugle.

En sortie de traitement une discrimination des performances de l'algorithme est réalisée sur la base de variables internes au traitement afin d'écarter les échantillons trop complexes ou dégradés pour être automatisés de manière fiable.

*Distribution du nombre d'audios du set en fonction du MCL prédit : données Elfe*



## Distribution du nombre d'audios du set en fonction du MCLM prédit : données Epipage2



Dans une seconde phase d'analyse des données, des points de donnée où la performance du modèle est faible ont été détectés. Ces points requièrent un traitement manuel. Cette discrimination est réalisée à l'aide de l'algorithme de Machine Learning *Isolation Forest*.

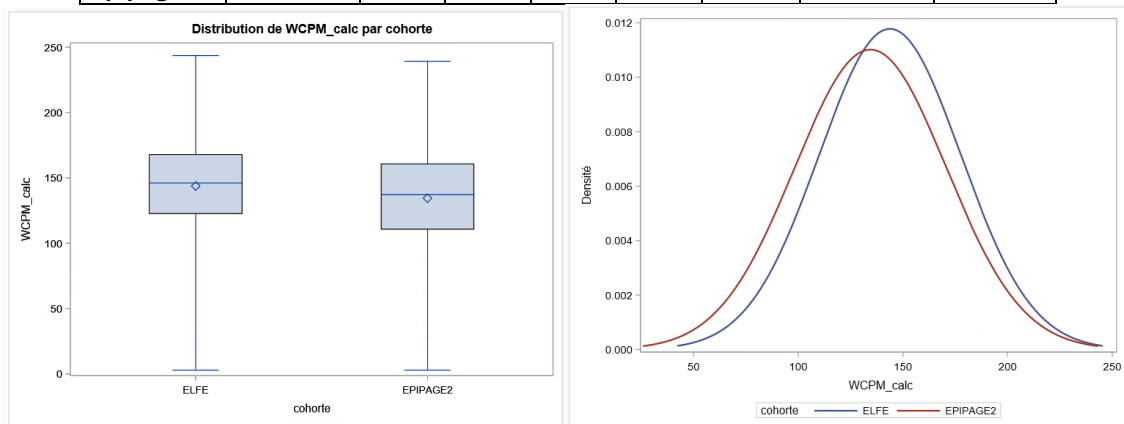
Ont été identifiés : pour Elfe, **146 points de donnée, soit 1,9% du set total** ; Pour EPIPAGE2, **70 points de donnée, soit 4,7% du set total**.

➔ En ne considérant que la modalité "outlier" = 1 (recodage par I.A.)

### 3. Description des variables pour les échantillons fiables

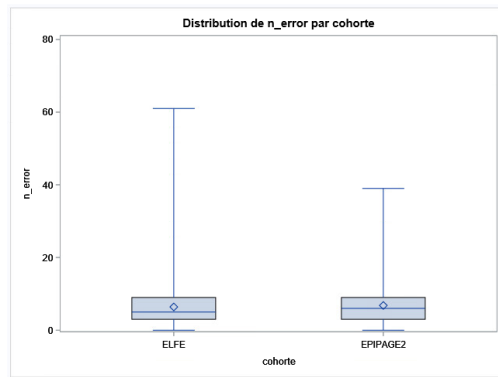
#### Estimation du nombre de mots correctement lus par minute

	N	Min	q25	q50	q75	Max	Moyenne	Ec type
<b>Elfe</b>	7372	2.9	122.7	146.1	167.9	243.7	<b>143.81</b>	33.86
<b>Epipage2</b>	1421	2.9	110.8	137.3	160.7	239.3	<b>134.48</b>	36.21



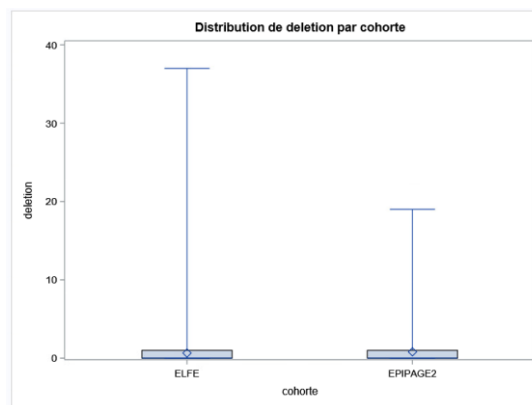
#### Nombre d'erreurs total

	N	Min	q25	q50	q75	Max	Moyenne	Ec type
<b>Elfe</b>	7372	0,00	3,00	5,00	9,00	61,00	<b>6,41</b>	4,88
<b>Epipage2</b>	1421	0,00	3,00	6,00	9,00	39,00	<b>6,81</b>	4,95



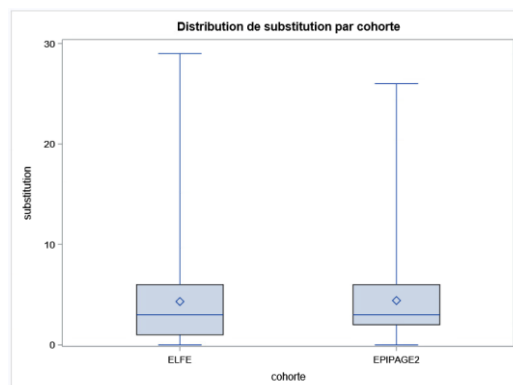
### Nombre d'erreurs de type suppression

	N	Min	q25	q50	q75	Max	Moyenne	Ec type
<b>Elfe</b>	7372	0,00	0,00	0,00	1,00	37,00	<b>0.65</b>	1.52
<b>Epipage2</b>	1421	0,00	0,00	0,00	1,00	19,00	<b>0.81</b>	1.68



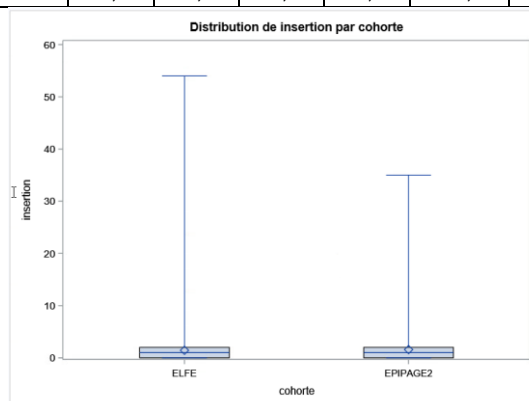
### Nombre d'erreurs de type substitution

	N	Min	q25	q50	q75	Max	Moyenne	Ec type
<b>Elfe</b>	7372	0,00	1,00	3,00	6,00	29,00	<b>4.32</b>	3.93
<b>Epipage2</b>	1421	0,00	2,00	3,00	6,00	26,00	<b>4.42</b>	3.84



## Nombre d'erreurs de type insertion

	N	Min	q25	q50	q75	Max	Moyenne	Ec type
Elfe	7372	0,00	0,00	1,00	2,00	54,00	1.44	1.85
Epipage2	1421	0,00	0,00	1,00	2,00	35,00	1.58	2.07



## Complément 3 - Éléments relatifs à la conversion des notes brutes en notes standardisées

### Pour les matrices, les puzzles et le PPVT, notes standardisées sur l'âge

La standardisation selon l'âge permet de tenir compte des âges des enfants et de comparer les résultats obtenus dans Elfe et Epipage à ceux obtenus avec les mêmes outils dans d'autres études. Le tableau suivant présente les éléments repris du manuel WISC-V (2014) pour les matrices et les puzzles, du manuel PPVT (2019) et pour le test de fluence, la standardisation se fait selon le guide de passation EL.F.E de Cognisciences (Juton et Lequette, 2013) en considérant qu'en France, un enfant de 10,5 ans est scolarisé en CM2 quand il n'a pas d'avance ou de retard scolaire.

Test	Quartiles observés dans Elfe -> Notes standardisées	Quartiles observés dans EPIPAGE-> Notes standardisées
Matrices à l'âge de 10 ans 4 mois – 10 ans 7 mois	<b>Q2 : 19 -&gt; 11</b> Q1 : 16 -> 8 Q3 : 20 -> 12	<b>Q2 : 17-&gt; 9</b> Q1 : 15 -> 7 Q3 : 19 -> 11
Puzzles à l'âge de 10 ans 4 mois – 10 ans 7 mois	<b>Q2 : 17 -&gt; 11</b> Q1 : 15 -> 9 Q3 : 19 -> 12	<b>Q2 : 16 -&gt; 10</b> Q1 : 13 -> 8 Q3 : 18 -> 11
PPVT à l'âge de 10 ans 4 mois – 10 ans 7 mois	<b>Q2 : 179 -&gt; 100</b> Q1 : 173 -> 93 Q3 : 184 -> 108	<b>Q2 : 179 -&gt; 100</b> Q1 : 173 -> 93 Q3 : 184 -> 108
Mr Petit en classe de CM2	<b>Q2 : 146.1 -&gt; p50</b> Q1 : 122.7 -> p30 Q3 : 167.9 -> p80	<b>Q2 : 137.3 -&gt; p40</b> Q1 : 110.8 -> p15 Q3 : 160.7 -> p70

### Étalonnage Mr Petit

La figure suivante fournit l'étalonnage du CE1 à la 5<sup>e</sup> de la variable MCLM de Mr Petit. Les seuils sont utilisés pour établir les modalités de la variable MPETIT\_MCLM\_STD. Les seuils p15 et p5 permettent d'identifier les enfants ayant un score normal de fluence à l'épreuve des enfants ayant un retard en fluence et de ceux présentant des difficultés importantes en lecture. Ils sont utilisés pour catégoriser le score de l'enfant selon la classe de scolarisation de référence en CM2 en trois modalités (normal, inférieur, difficultés importantes en lecture) variable MPETIT\_MCLM\_STD mise à disposition dans la base Elfe.

Figure 9 : Étalonnage du CE1 à la 5<sup>e</sup> de Mr Petit

	CE1 Moyenne 68 Ecart-type 28	CE2 Moyenne 95 Ecart-type 26	CM1 Moyenne 116 Ecart-type 33	CM2 Moyenne 137 Ecart-type 31	6 <sup>e</sup> Moyenne 141 Ecart-type 32	5 <sup>e</sup> Moyenne 154 Ecart-type 32
p90	103	129	161	177	181	198
p80	90	116	144	164	164	180
p70	79	108	131	154	157	170
p60	70	101	124	147	149	163
p50	64	95	115	140	142	153
p40	57	88	109	125	132	144
p30	52	80	98	117	123	138
p25	20	78	93	114	121	135
p20	48	73	87	112	117	125
p15	41	69	80	108	106	122
p10	37	62	74	96	99	115
<b>p5</b>	24	54	67	87	86	104

Note de lecture. Selon les auteurs, le percentile 15 indique les enfants présentent un retard de fluence en lecture et le percentile 5 ceux présentant des difficultés très importantes de lecture. Un score supérieur à p30 est considéré comme n'handicapant pas la compréhension lors de la lecture.