

Évaluation informatisée du vocabulaire chez les enfants de 8 à 11 ans

Computer-Based Vocabulary Assessment in Children Aged 8 to 11

Jean Écalle, Université Lyon 2

Nicolas Bailloud, Université Lyon 2

Émilie Dujardin, Université Lyon 2

Annie Magnan, Université Lyon 2

Résumé

Cette étude pilote a pour objectif de construire une nouvelle épreuve informatisée d'évaluation du vocabulaire en langue Française, destinée aux élèves au primaire. Des tâches à choix multiples mesurant l'étendue (i.e., Image) et la profondeur du vocabulaire (i.e., Synonyme, Définition) ont été administrées à 255 enfants. La rapidité et l'exactitude des réponses ont été enregistrées. Comme attendu, les analyses ont révélé que le niveau scolaire, le genre, et la fréquence lexicale des items impactaient les temps de réponse et/ou les réponses correctes. Les performances s'améliorent avec le niveau scolaire et pour les mots fréquents. Les filles semblent avoir un vocabulaire plus riche que celui des garçons (e.g., profondeur) et elles semblent moins sensibles à la fréquence des mots que les garçons. Une analyse des qualités psychométriques de l'épreuve a permis de retenir une épreuve en 36 items ayant une consistance interne satisfaisante. Enfin, une analyse factorielle confirmatoire n'a pas permis de mettre en évidence une structure en deux facteurs du vocabulaire (e.g., profondeur, étendue). Les résultats sont discutés en termes de qualité de représentations lexicales et de choix des tâches.

Mots-clés : vocabulaire ; évaluation ; outil numérique ; enfants

Abstract

The purpose of this pilot study is to construct a new computerized French language vocabulary assessment test for elementary school children. Multiple choice tasks measuring vocabulary breadth (e.g., image) and depth (e.g., synonym, definition) were administered to 255 children. The speed and accuracy of responses were recorded. As expected, analyses revealed that grade level, gender, and lexical frequency of items impacted response times and/or correct responses. Performance improved with grade level and for frequent words. Girls appear to have a richer vocabulary than boys (e.g., depth) and they appear to be less sensitive to word frequency than boys.

An analysis of the psychometric qualities of the test allowed us to retain a 36-item test with satisfactory internal consistency. Finally, a confirmatory factor analysis did not reveal a two-factor structure of vocabulary (e.g., depth, breadth). The results are discussed in terms of the quality of lexical representations and task choice.

Keywords: vocabulary; assessment; digital tool; children

Introduction

Le vocabulaire constitue une composante du langage essentielle dans les acquisitions scolaires. Plus précisément, à maintes reprises il a été vu comme un puissant prédicteur de la compréhension en lecture. Toutefois, la question de son évaluation continue de se poser à plusieurs titres. D'une part, les dispositifs, le matériel linguistique et les épreuves utilisées sont nombreux et variés, ce qui amène à la question de savoir quels outils peuvent être utilisés. D'autre part, à l'heure du numérique, cette question revient d'autant plus à l'ordre du jour pour imaginer et construire de nouvelles épreuves numériques. Il s'agira aussi d'examiner l'effet potentiel des caractéristiques individuelles et lexicales sur les performances de ces épreuves. Il faut alors tester et évaluer les qualités psychométriques de ces nouveaux outils informatisés. C'est sur cette problématique, centrale dans le nouveau contexte des outils numériques disponibles à l'école, que s'est focalisée cette étude menée auprès d'enfants du primaire. L'objectif de notre article est d'examiner le rôle des variables individuelles (p. ex., genre, âge) et lexicales (p. ex., fréquence lexicale) sur le vocabulaire réceptif à l'aide d'un nouveau test numérique. Par ailleurs, nous examinons les qualités psychométriques de notre test afin de déterminer sa fiabilité et sa sensibilité. Dans la partie introductive, nous évoquerons l'importance d'évaluer le vocabulaire, en précisant des tests spécifiques pour chacune des dimensions du vocabulaire (c.-à-d., tests papier-crayon et/ou informatisés). Puis, nous définirons certaines caractéristiques liées aux individus (p. ex., genre, âge) ou aux mots (p. ex., fréquence lexicale) pouvant influencer le niveau de vocabulaire.

Évaluer le vocabulaire : des tests papier-crayon aux tests numériques

Le vocabulaire est un concept multidimensionnel complexe dont les liens avec la lecture ont été largement étudiés. Pourtant, son rôle précis dans les modèles de lecture reste encore indéterminé. Les liens entre les connaissances du vocabulaire, qu'elles soient liées à l'étendue ou à la profondeur, et les capacités de compréhension de lecture, orale et de la lecture de mot ont été à plusieurs reprises examinées sans pour autant arriver à un consensus (voir Magnan & Écalle, 2018; Oakhill *et al.*, 2019 pour des revues). De plus, alors que la plupart des tests de vocabulaire ont des formats papier/crayon, des chercheurs ont montré l'avantage de l'utilisation d'outil numérique (voir Neumann & Neumann, 2019). Peu d'outils informatiques sont actuellement disponibles pour évaluer le vocabulaire. Pourtant, les technologies numériques présentent de multiples avantages pour l'évaluation, tels qu'une validité et une fiabilité accrues, un engagement plus important dans la tâche par les personnes testées, une mise en œuvre plus facile en classe, la possibilité d'enregistrer les stratégies et les mesures de performance des personnes testées. En effet, outre le côté ludique de la version numérique d'un test, l'avantage réside dans une standardisation des mesures et une rapidité des passations. Cela permet d'enregistrer de façon précise la rapidité de réponse (c.-à-d., temps de

réponse) et l'exactitude des réponses (c.-à-d., réponse correcte). Ces deux mesures rendent compte de la qualité d'accès aux connaissances en vocabulaire des enfants (Perfetti, 2007; Richter *et al.*, 2013).

Essayer de préciser le rôle des deux dimensions du vocabulaire implique de les mesurer de façon distincte. Sans prétendre à l'exhaustivité, nous présentons les principales épreuves utilisées sur un support papier-crayon. Les mesures de vocabulaire sont souvent des mesures du vocabulaire réceptif (c.-à-d., les mots que l'individu comprend) plutôt que productif (c.-à-d., les mots que l'individu peut produire, utiliser) (voir Pearson *et al.*, 2007 pour une revue). Tout d'abord, le vocabulaire réceptif est souvent plus large que le vocabulaire productif, c'est-à-dire qu'un individu comprend plus de mots qu'il n'en utilise (Pearson *et al.*, 2007). Ensuite, le vocabulaire réceptif est facilement mesurable, à l'aide de réponses à choix multiples par exemple.

L'étendue du vocabulaire réceptif est traditionnellement évaluée au moyen de tests sur papier utilisant un paradigme d'images à quatre choix dans lequel les enfants sont invités à faire correspondre un mot cible prononcé à voix haute par l'expérimentateur à l'une des quatre images. La tâche est facile à faire passer par le praticien/expérimentateur et aussi facile à comprendre par l'enfant. Le test de vocabulaire réceptif le plus largement utilisé et le plus connu est l'Échelle de vocabulaire en images (*Peabody Pictures Vocabulary Test* - en anglais américain, PPVT-4, Dunn & Dunn, 2007), qui a été adapté pour être utilisé dans de nombreuses autres langues, telles que l'anglais britannique (*British Picture Vocabulary Scale*, Dunn *et al.*, 2009) et le français (Dunn *et al.*, 1993). Ce test permet de déterminer le niveau réceptif du langage de l'enfant et les difficultés d'apprentissage.

Mesurer la profondeur du vocabulaire est plus difficile, car elle est reliée aux connaissances que l'individu a d'un mot, son sens général, son sens spécifique selon le contexte, mais aussi ses synonymes ou ses antonymes. Elle est souvent mesurée avec un test de vocabulaire expressif. Dans ce type de test, l'individu doit fournir par exemple une définition d'un mot cible. Les sous-test « Vocabulaire expressif » de l'échelle d'intelligence pour enfants (*Wechsler Intelligence Scale for Children* – WISC; Wechsler, 2003 en anglais; Wechsler, 2005 en français) et du test de connaissance de mots (*Test of Word Knowledge* – TOWK, Wiig & Secord, 1992 en anglais) sont souvent utilisés dans la littérature. Le sous-test « vocabulaire » du DIBELS (*Dynamic Indicators of Basic Early Literacy Skills*, en anglais, Good & Kaminski, 2002; en français, Dufour-Martin & Good, 2010) propose d'évaluer la profondeur du vocabulaire en demandant aux enfants d'utiliser correctement les mots dans les phrases. Produire une définition d'un mot ou utiliser correctement un mot dans une phrase peut être difficile pour certaines populations (p. ex., jeunes enfants, enfants au développement atypique, etc.). Certains tests de la profondeur du vocabulaire réceptif ont été proposés, dans lesquels l'individu doit indiquer si des mots partagent des connexions sémantiques (p. ex., *Clinical Evaluation of Language Fundamentals, Fifth Edition* - CEFL-5; Wiig *et al.*, 2013 en anglais; Wiig *et al.*, 2009 en français canadien). Le TOWK (Wiig & Secord, 1992) inclut également un test dans lequel les participants doivent choisir la signification d'expressions dans un format à choix multiple ou sélectionner le synonyme d'un mot cible parmi quatre distracteurs. Les mesures des connaissances du vocabulaire sont nombreuses, avec des formats de présentations différents.

L'intérêt croissant pour les outils numériques a permis de développer des applications. Différentes batteries de tests ont d'ailleurs été implémentées sur ordinateur (p. ex., CEFL, WISC,

DIBELS). L'avantage des tablettes avec la simplification de leur utilisation (pas de clavier à gérer) les rend accessibles même aux enfants ayant des difficultés de langage (par exemple, voir Marble-Marble-Flint *et al.*, 2019 pour une étude sur les enfants atteints de troubles du spectre autistique). L'évaluation sur tablette du vocabulaire réceptif est plus facile à mettre en place, car contrairement à un test expressif, qui nécessite un logiciel de reconnaissance vocale précis et souvent très long à développer. Des épreuves de l'évaluation de l'étendue du vocabulaire avec un paradigme de choix multiples de l'image chez les enfants multilingues (germano-turc, Schaefer *et al.*, 2019), ainsi que chez les enfants monolingues anglophones (Schaefer *et al.*, 2016) ont été informatisées. Toutefois, des épreuves informatisées mesurant la profondeur du vocabulaire ne semblent pas encore disponibles et devraient encore être implémentées. La batterie de test DiCoLec (diagnostic de la compréhension en lecture) a été utilisée pour évaluer la lecture et la compréhension orale, la lecture des mots et le vocabulaire (Auphan *et al.*, 2019, 2020). Le sous-test associé à l'évaluation de la profondeur vocabulaire proposait des questions sur les attributs des mots cibles. Les enfants devaient indiquer si le mot cible possédait l'attribut ou non. Les nouvelles technologies permettent ainsi l'administration standardisée des tâches et, plus important encore, la collecte automatisée et plus rapide d'une série de données, de types de réponses (réponses correctes et erreurs) et de temps de réponse (Beauvais *et al.*, 2018; Frank *et al.*, 2016; Neumann & Neumann, 2019; Richter *et al.*, 2013). Il est maintenant généralement reconnu que deux paramètres, à savoir la rapidité et la précision, sont importants lors de l'évaluation des compétences linguistiques (p. ex., Richter *et al.*, 2013).

Impact de variables individuelles et lexicales sur les performances en vocabulaire

Le niveau de vocabulaire est impacté par des caractéristiques individuelles (genre, âge; Dunn & Dunn, 1981, 1997; Fenson *et al.*, 2007; Rice & Hoffman, 2015; Simos *et al.*, 2011; Taylor *et al.*, 2013) ou encore par des caractéristiques lexicales (fréquence lexicale; Schaefer *et al.*, 2016).

Une récente étude longitudinale sur le long terme, menée auprès de participants âgés entre 2,6 et 21 ans, a montré un effet de l'âge sur la croissance du vocabulaire (Rice & Hoffman, 2015). Le taux d'acquisition du vocabulaire augmentant jusqu'à 12 ans avant de ralentir. Des facteurs environnementaux jouent également un rôle sur l'acquisition du vocabulaire, comme le milieu familial, le niveau d'étude de la mère, ou encore l'exposition à la lecture (p. ex., Hoff, 2006; Sénéchal *et al.*, 2017). Le vocabulaire et les capacités de lecture ont une relation bidirectionnelle. En effet, plus les individus connaissent de mots, meilleure est leur compréhension de la lecture; et plus ils lisent, plus leur vocabulaire s'enrichit (p. ex., Oakhill *et al.*, 2019).

Il existe des différences dans l'acquisition des premiers mots de l'enfant en fonction du genre, avec un avantage pour les filles (Zubrick *et al.*, 2007; mais voir Simos *et al.*, 2011). Toutefois, les travaux sur les différences de niveau de vocabulaire entre les filles et les garçons en faveur des premières n'amènent pas à un consensus clair. L'effet de genre devrait être examiné dans les études de façon systématique, notamment pour savoir si des normes des tests en fonction du genre devraient être fournies. Alors que Dunn et Dunn (1981) ont constaté que les garçons obtenaient de meilleurs résultats au PPVT que les filles, Fenson *et al.* (2007) ont montré l'inverse : ce sont les filles qui avaient un vocabulaire plus étendu que les garçons. Rice et Hoffman (2015) ont montré un effet du genre sur la croissance du vocabulaire. Au départ, la croissance du vocabulaire était plus importante chez les jeunes filles, puis elle s'est ensuite stabilisée avec l'âge avant d'être plus importante pour les

garçons de 10 à 21 ans (Rice & Hoffman, 2015). Au cours du développement, la croissance du vocabulaire est influencée en fonction du genre. Des chercheurs avancent une explication d'ordre hormonal ou génétique (voir Hoff, 2006; Rice & Hoffman, 2015). Toutefois, nous pouvons aussi supposer que l'intérêt des enfants pour le langage pourrait aussi expliquer ces différences (par exemple, l'intérêt pour la lecture varie en fonction du sexe et/ou de l'âge) (voir Hoff, 2006).

Nous avons évoqué des facteurs liés à l'individu, mais un autre facteur affecte les performances de vocabulaire lors des évaluations, et il est important de tenir compte des caractéristiques des mots, et en particulier de leur fréquence, lors de l'examen du vocabulaire (Schaefer *et al.*, 2016). La fréquence des mots est mesurée à partir du nombre d'occurrences avec laquelle un mot apparaît dans la vie quotidienne d'un individu (livres, journaux, films). Plus un mot est fréquent, mieux il est connu et, par conséquent, plus il est susceptible de faire partie du vocabulaire des enfants. La connaissance des mots dépend également du temps de lecture : plus les enfants sont exposés à la lecture, y compris aux activités de lecture partagée de livres chez les plus jeunes, plus leur vocabulaire s'accroîtra et plus ils seront capables de redéfinir les mots qu'ils connaissent déjà (par exemple, Cunningham, 2005; Oakhill *et al.*, 2019). Ces différentes variables feront l'objet d'analyses dans l'étude que nous présentons ici.

Questions de recherche

L'objectif de cette étude est de construire une nouvelle épreuve destinée aux élèves de primaire pour évaluer leur niveau de vocabulaire, épreuve avec deux caractéristiques essentielles, l'une en la proposant dans un format numérique (sur le Web) pour obtenir les deux paramètres, précision (qualité des réponses) et vitesse (temps pour accéder au sens des mots), la seconde en distinguant trois types de tâches différentes pour évaluer l'étendue du vocabulaire et la profondeur. Nous avons déterminé des facteurs liés aux caractéristiques individuelles (niveau scolaire et genre) et aux caractéristiques lexicales (fréquence) impactant les performances. Nous nous attendons à ce que les performances progressent (les réponses correctes augmentant et les temps de réponse diminuant) avec le niveau scolaire (donc l'âge), qu'elles soient supérieures chez les filles et pour les mots les plus fréquents avec des temps de réponse inférieurs. Par ailleurs, la nouveauté de l'épreuve nous conduira à examiner la qualité psychométrique des items au cours de diverses analyses pour sélectionner, in fine, ceux qui contribuent positivement au score total. Enfin, nous testerons à l'aide d'analyses appropriées (modèles d'équations structurales) si les épreuves rendent bien compte (ou non?) des deux composantes, étendue et profondeur, du vocabulaire.

Méthode

Participants

À partir d'une base de données recueillies auprès d'environ 300 élèves répartis de la deuxième année de primaire (G2)¹ à la quatrième année (G5), les données finales ont été examinées

¹ Cette abréviation renvoie aux grades propres au cursus des écoles des pays anglophones.

sur un effectif de 255 (136 garçons/119 filles) après avoir écarté celles qui n'étaient pas exploitables (incident technique ou évaluation inachevée) et celles des élèves ayant au moins un an de retard scolaire. Finalement, les analyses ont porté sur quatre classes, 50 élèves de G2 (âge moyen = 7,8 ans; é-t = 3,7 mois), 73 G3 (am = 8,9 a; é-t = 3,3 m), 70 G4 (am = 9,7 a; é-t = 4.4) et 62 G5 (10,7 a; é-t = 4 m). L'administration des épreuves s'est déroulée de janvier à mars 2018 après avoir obtenu l'accord des enseignants, celui des parents et les autorisations académiques.

Matériel et procédure

Trois épreuves de vocabulaire ont été proposées dans une salle informatique de l'établissement, les enfants étant munis de casque sur les oreilles. Les items (85 au total) étaient présentés en double modalité, orale-écrite, afin d'éviter d'éventuelles difficultés de lecture pour certains enfants.

Image

Cette épreuve de type classique consistait à choisir l'image parmi quatre qui correspond au mot présenté au centre de l'écran et entendu en même temps. Les images apparaissaient simultanément et l'enfant devait cliquer à l'aide de la souris sur l'image correspondant à l'item (Figure 1a)². Quatre items tests étaient présentés, l'item cible, deux items du même champ sémantique et un intrus. Dix verbes et dix noms de fréquence lexicale différente étaient proposés dans un ordre aléatoire pour chaque enfant.

Synonyme

L'enfant devait retrouver le mot cible (*peau*) proche du mot source (*écorce*) (Figure 1b) parmi quatre items tests, le mot cible, deux mots du même champ sémantique et un intrus. Dix noms et dix adjectifs variant sur leur fréquence lexicale étaient à nouveau proposés dans un ordre aléatoire.

Définitions

Après la présentation d'un mot source (Figure 1c), trois définitions étaient présentées successivement à l'écran et dans le casque. L'enfant devait cliquer sur le mot correspondant à sa définition. Deux items distracteurs, définition proche ou contrastée par rapport à celle du mot source, étaient proposés. Quinze noms, 15 verbes et 15 adjectifs étaient présentés dans un ordre aléatoire.

Les trois épreuves ont été administrées dans un ordre aléatoire avec une pause entre chaque. L'ensemble de la passation a duré environ 45 minutes. Le temps pour répondre n'était pas contraint, mais la consigne donnée avant la passation des épreuves insistait sur un temps de réponse aussi rapide que possible tout en précisant qu'il fallait donner une réponse correcte. Pour les trois tâches le temps était déclenché dès que tous les items étaient présentés à l'écran (et dans le casque). Toutes les réponses correctes (Rc) et fausses, et les temps de réponses (TR) ont été recueillis sur un serveur Web puis transposés dans un tableur Excel. L'ensemble des données a donné lieu à un certain nombre de tris d'abord, puis d'analyses ensuite.

² Le développement des tâches et le dispositif pour saisir les données ont été réalisés par Florian Chassaing, médiateur co-design et développeur à Canopé (Lyon).

Figure 1

Captures d'écran dans les tâches Image (a), Synonyme (b), et Définition (c)



Résultats

Examen des données brutes.

L'objectif est d'une part, de repérer les items qui n'ont pas donné lieu à un taux de réussite supérieur au hasard, puis de lisser les TR. Pour les items, dans les tâches « Image » et « Synonyme », le taux de réussite proche du hasard est de 25 % et celui dans la tâche « Définition » est de 33 %. Les items dont les scores ne se distinguaient pas du hasard ($p < ,05$) ont été écartés des analyses suivantes ce qui correspond à un total de 25 items sur 85 (voir Laveault & Grégoire, 2014). Le tableau 1 présente le nombre d'items retenus par catégorie lexicale et selon deux niveaux de fréquence contrastés en utilisant l'indice SFI de Manulex (voir Lété, Sprenger-Charolles & Colé, 2004)³; on observe que tous les items de la tâche Image sont maintenus.

Nous avons ensuite procédé à un lissage des TR (relevés en centième de seconde) obtenus sur les Rc, lissage effectué en trois étapes. D'abord nous avons enlevé tous les TR inférieurs à 0,25 seconde (ces temps trop rapides, très rares semblent provenir d'un problème technique) et tous

³ L'indice SFI de la base Manulex (voir <http://www.manulex.org/fr/home.html>) signifie « Index de fréquence standard » (en français).

ceux supérieurs à 10 secondes. En effet, il nous semble que si le mot est moins précisément connu ce qui peut donner lieu à quelques hésitations, le seuil de 10 secondes est largement suffisant pour fournir une réponse. Ainsi, nous avons écarté 17 % des TR sur l'ensemble des TR avec Rc. La seconde étape a consisté à calculer la moyenne des TR par enfant toute tâche confondue. Enfin, nous avons remplacé les TR en dehors des limites par la moyenne précédemment calculée, ceci sur les trois tâches et pour chaque enfant.

Tableau 1

Nombre d'items retenus dans chaque tâche en fonction de leur catégorie lexicale et de leur fréquence

| | | F1 | | | F2 | | | TotRet |
|------------|-----|-----|-------------|-------|-----|-------------|-------|--------|
| | | Nom | Adjectif | Verbe | Nom | Adjectif | Verbe | |
| Image | SFI | | 64,1 – 48,6 | | | 46,4 – 19,4 | | |
| | nb | 5 | - | 5 | 5 | - | 5 | 20 |
| Synonyme | SFI | | 61,7 – 45,8 | | | 44,6 – 22,7 | | |
| | nb | 4 | 3 | - | 4 | 3 | - | 14 |
| Définition | SFI | | 58,7 – 36,6 | | | 35,1 - 15 | | |
| | nb | 1 | 7 | 5 | 4 | 5 | 4 | 26 |

Effet du niveau scolaire, du genre et de la fréquence lexicale

Les statistiques descriptives des 60 items retenus sont présentées dans le tableau 2 pour les Rc et le tableau 3 pour les TR. Une série de Manovas, avec les facteurs inter, Classe (G2, G3, G4, G5) et Genre (fille vs garçon) et le facteur intra Fréquence lexicale (F1, F2) a été conduite pour chaque tâche et sur les deux variables, Rc et TR. Une synthèse des principaux résultats issus de ces analyses est fournie dans deux tableaux présentant seulement les effets significatifs des facteurs principaux.

Tableau 2

Moyennes (écart-types) des réponses correctes dans les trois tâches (n = 255) en fonction des niveaux scolaires, du genre et de la fréquence lexicale

| | | G2 | | G3 | | G4 | | G5 | |
|-----|----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | | f | g | f | g | f | g | f | g |
| | | (n = 22) | (n = 28) | (n = 32) | (n = 41) | (n = 32) | (n = 38) | (n = 33) | (n = 29) |
| Déf | F1 | 6,82 (2,26) | 5,61 (1,93) | 6,88 (2,6) | 6,63 (2,32) | 8 (2) | 7,71 (2,26) | 9,24 (1,9) | 8,1 (2,14) |
| | F2 | 5 (2,07) | 4,82 (2,28) | 6,31 (2,46) | 5,83 (2,21) | 7,13 (2,2) | 6,97 (2,07) | 8,21 (2,15) | 7,41 (2,01) |
| Ima | F1 | 7,96 (1,09) | 7,82 (1,22) | 8,47 (1,34) | 8,15 (1,3) | 8,75 (1,32) | 8,32 (1,56) | 8,88 (,86) | 8,97 (,82) |
| | F2 | 6,09 (1,69) | 6,18 (1,81) | 6,28 (1,76) | 6,12 (1,98) | 6,94 (1,64) | 6,71 (2,22) | 7,76 (1,3) | 7,31 (1,31) |
| Syn | F1 | 3,23 (1,57) | 3,07 (1,36) | 3,72 (1,49) | 3,83 (1,43) | 4,63 (1,43) | 4,08 (1,15) | 4,91 (1,07) | 4,52 (1,06) |
| | F2 | 2,82 (1,01) | 2,61 (1,2) | 2,75 (1,19) | 2,95 (,97) | 3,13 (1,56) | 2,97 (1,15) | 3,15 (1,12) | 3,31 (,76) |

Notes : Déf : Définition (max = 13 par condition); Ima : Image (max = 10 par condition); Syn : Synonyme (max = 7 par condition).

Tableau 3

Moyennes (écart-types) des temps de réponses lissés (en seconde) dans les trois tâches (n = 255) en fonction des niveaux scolaires, du genre et de la fréquence lexicale

| | | G2 | | G3 | | G4 | | G5 | |
|-----|----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | | f ^a | g | f | g | f | g | f | g |
| | | (n = 22) | (n = 28) | (n = 32) | (n = 41) | (n = 32) | (n = 38) | (n = 33) | (n = 29) |
| Déf | F1 | 4,38 (1,29) | 3,61 (0,88) | 3,94 (1,48) | 3,47 (1,28) | 4,45 (1,24) | 4,1 (1,14) | 4,19 (1,17) | 3,68 (1,06) |
| | F2 | 3,99 (1,09) | 3,75 (1,63) | 3,94 (1,26) | 3,77 (1,38) | 4,29 (1,1) | 3,92 (1,07) | 3,93 (1,01) | 3,63 (1,31) |
| Ima | F1 | 3,98 (1,26) | 3,48 (1,04) | 3,4 (1,2) | 3,42 (0,97) | 3,14 (0,99) | 3,2 (1,12) | 2,88 (1,21) | 3 (1,2) |
| | F2 | 4,03 (1,27) | 4,16 (1,18) | 3,56 (1,05) | 4 (1,29) | 3,46 (1,23) | 3,67 (1,1) | 3,01 (1,07) | 3,31 (1,15) |
| Syn | F1 | 4,18 (1,83) | 3,91 (1,4) | 3,93 (1,46) | 3,91 (1,72) | 4,03 (1,55) | 3,69 (1,49) | 3,51 (1,12) | 3,48 (1,34) |
| | F2 | 4,22 (1,72) | 3,42 (1,64) | 3,51 (1,45) | 3,5 (1,24) | 3,22 (1,2) | 3,52 (1,04) | 3,27 (1,57) | 3,1 (1,14) |

Notes : ^a certaines moyennes sont calculées sur un effectif réduit d'une ou deux unités en raison d'une absence de TR < 10 sec pour tous les items d'une seule condition; Déf : Définition (max = 13 par condition); Ima : Image (max = 10 par condition); Syn : Synonyme (max = 7 par condition).

Pour les Rc (Tableau 4), nous observons un effet du facteur Classe dans les trois tâches, les scores augmentant de G2 à G5. Un effet du facteur Genre est relevé dans la tâche « Définition » en faveur des filles. Enfin, un effet de Fréquence est également observé dans les trois tâches, les scores étant supérieurs pour les mots plus fréquents.

Tableau 4

Synthèse des effets principaux issus des manovas sur les variables réponses correctes (Rc) pour les trois tâches

| | | Classe | Genre | Fréquence |
|------------|------------|---------|-------|-----------|
| Définition | F | 23,31 | 5,76 | 36,09 |
| | p | ,0001 | ,02 | ,0001 |
| | η^2_p | ,22 | ,02 | ,13 |
| | moy | G2=5,56 | G=6,6 | F1=7,37 |
| | | G3=6,41 | F=7,2 | F2=6,46 |
| | G4=7,45 | | | |
| | G5=8,24 | | | |
| Image | F | 10,36 | | 265,48 |
| | p | ,0001 | | ,0001 |
| | η^2_p | ,11 | | ,52 |

| | | Classe | Genre | Fréquence |
|----------|------------|---------|-------|-----------|
| moy | | G2=7,01 | | F1=8,41 |
| | | G3=7,25 | | F2=6,67 |
| | | G4=7,68 | | |
| | | G5=8,23 | | |
| Synonyme | F | 11,54 | | 118,22 |
| | p | ,0001 | | ,0001 |
| | η^2_p | ,12 | | ,32 |
| | moy | G2=2,93 | | F1=4 |
| | | G3=3,31 | | F2=2,96 |
| | | G4=3,7 | | |
| | | G5=3,97 | | |

Pour les TR (Tableau 5), l'effet attendu du facteur Classe est présent uniquement dans la tâche « Image », les TR décroissant de G2 à G5. L'effet du facteur Genre est relevé dans la tâche « Définition », les filles ayant des temps plus longs. Enfin, l'effet de Fréquence est observé dans la tâche « Image », les mots fréquents ayant un TR moyen plus rapide. En revanche dans la tâche « Synonyme », et de façon inattendue, l'effet de Fréquence est en faveur des mots moins fréquents, les TR pour F1 étant supérieurs aux TR pour F2.

Tableau 5

Synthèse des effets principaux issus des manovas sur les temps de réponse (TR en seconde) pour les trois tâches

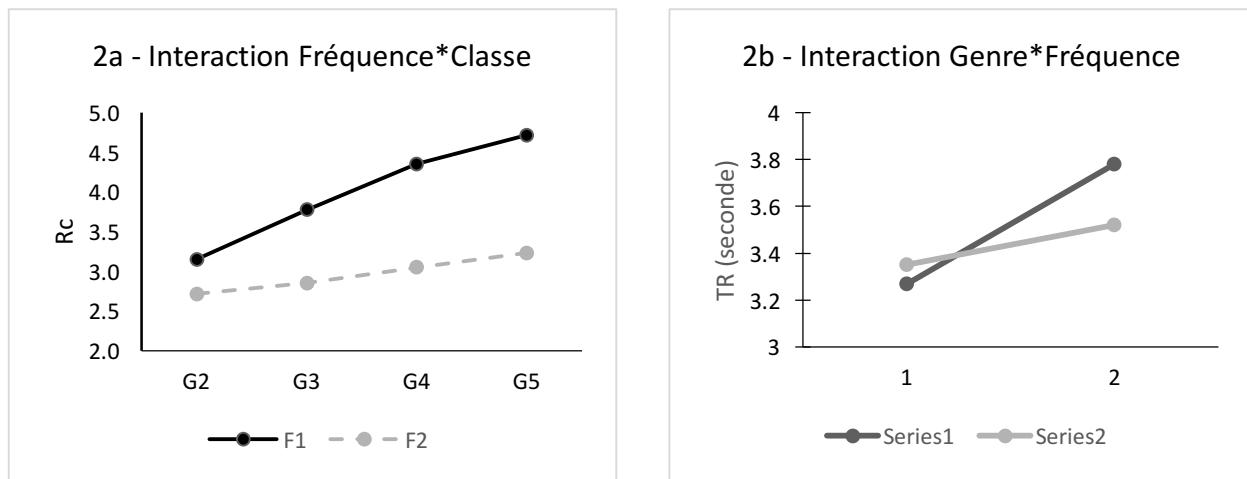
| | | Classe | Genre | Fréquence |
|------------|------------|---------|--------|-----------|
| Définition | F | | 7,94 | |
| | p | | ,005 | |
| | η^2_p | | ,03 | |
| | moy | | G=3,74 | |
| | | | F=4,14 | |
| Image | F | 6,78 | | 31,07 |
| | p | ,0002 | | ,0001 |
| | η^2_p | ,08 | | ,11 |
| | moy | G2=3,91 | | F1=3,31 |
| | | G3=3,59 | | F2=3,65 |
| | | G4=3,37 | | |
| | | G5=3,05 | | |
| Synonyme | F | | | 13,17 |
| | p | | | ,0003 |
| | η^2_p | | | ,05 |
| | moy | | | F1=3,81 |
| | | | | F2=3,48 |

En outre, à l'issue des manovas réalisées, deux interactions significatives ont été relevées. Pour la variable Rc, dans la tâche « Synonyme », une interaction Classe*Fréquence, $F(3,247) = 5,28$,

$p = ,002$, $\eta^2_p = ,06$, montre que la différence de performance entre les mots F1 et F2 augmente avec le niveau scolaire (Figure 2a). Pour la variable TR dans la tâche « Image », l'interaction Genre*Fréquence, $F(1,245) = 7,95$, $p = ,005$, $\eta^2_p = ,03$, montre que pour les mots moins fréquents (F2), les filles répondent significativement plus vite que les garçons ($p = ,03$; post-hoc test Newman-Keuls; Figure 2b).

Figure 2

Graphiques des interactions significatives dans la tâche Synonyme (2a) et dans la tâche Image (2b)



Qualités psychométriques

Pour examiner les qualités psychométriques de cette épreuve vocabulaire, nous avons procédé en deux étapes. Nous avons d'abord sélectionné parmi les 60 items ceux qui contribuaient de façon non négligeable au score total (variable Rc). À partir de la matrice des r_{pbis} (coefficient de corrélation bisériale de point), nous avons rejeté les items dont le r_{pbis} était inférieur à ,20 (faible contribution). Sur les 36 items qui restent (Tableau 6), le coefficient α de Cronbach est de ,86 ce qui constitue un bon indice de consistance interne.

Tableau 6

Nombre d'items par catégorie

| Image | | Synonyme | Définition | |
|-------|----|----------|------------|----|
| F1 | F2 | | F1 | F2 |
| 6 | 6 | 8 | 8 | 8 |

Puis nous avons procédé à une analyse des items et de leurs caractéristiques à partir d'un modèle de réponse à l'item (MRI) (avec Stata 16). Plus précisément, nous avons utilisé un MRI à deux paramètres, l'un renvoie au niveau de difficulté de l'item (paramètre b) et le second à son pouvoir de discrimination (paramètre a). Nous présentons ici les courbes montrant les caractéristiques des items (*item characteristic curves*). Un graphique est représenté pour chaque tâche (Figures 3) avec les items contrastés⁴ sur le paramètre a . Les items à haut pouvoir de

⁴ La présentation de tous les items rendrait le graphique peu lisible.

discrimination sont ceux avec la pente la plus prononcée, soit les mots « serres » (graphique 3a), « anneau » (3b) et « incertitude » (3c) et les items les moins discriminants avec la pente plus douce pour les mots « mesurer » (3a), « narine » (3b) et « échouer » (3c) (Baker, 2001). Cinq des items à valeur négative de Θ correspondant au trait latent (ici le niveau en vocabulaire) indiquent des items plutôt faciles alors que l'item « échouer » (Figure 3b) est plus difficile.

Une ou deux composantes du vocabulaire? Les analyses ici basées sur les équations structurales ont pour objectif de répondre à la question : est-ce que les données Rc représentent un seul facteur, un niveau général de vocabulaire ou bien deux facteurs distinctifs que la littérature décrit comme l'étendue et la profondeur du vocabulaire? Pour y répondre deux analyses confirmatoires ont été réalisées avec le logiciel Mplus 8.1 (Muthén & Muthén, 2017) en utilisant l'indice d'estimation de maximum de vraisemblance robuste (MLR). Différents indices d'ajustement ont été examinés pour déterminer l'acceptabilité des modèles testés. Premièrement, le ratio entre le χ^2 et les degrés de liberté devrait être inférieurs à 3 (Schreiber *et al.*, 2006). Deuxièmement, les modèles acceptables ont des valeurs supérieures à ,90 pour le TLI (*Tucker-Lewis Index*) et le CFI (*Comparative Fit Index*) (Hu & Bentler, 1999), et des valeurs inférieures à ,08 pour le RMSEA (*Root Mean Square Error of Approximation*) et le SRMR (*Standardized Root Mean Square Residual*) (Schreiber *et al.*, 2006). Enfin, les plus petites valeurs de l'AIC (*Akaike Information Criterion*) et du BIC (*Bayes Information Criterion*) indiquent le modèle le plus parcimonieux (Burnham *et al.*, 2002; voir aussi Tabachnick & Fidell, 2012).

Figure 3

Courbes caractéristiques des items dans les tâches Image (a) (items : serres et mesurer), Synonyme (b) (items anneau et narine), et Définition (c) (items incertitude et échouer)

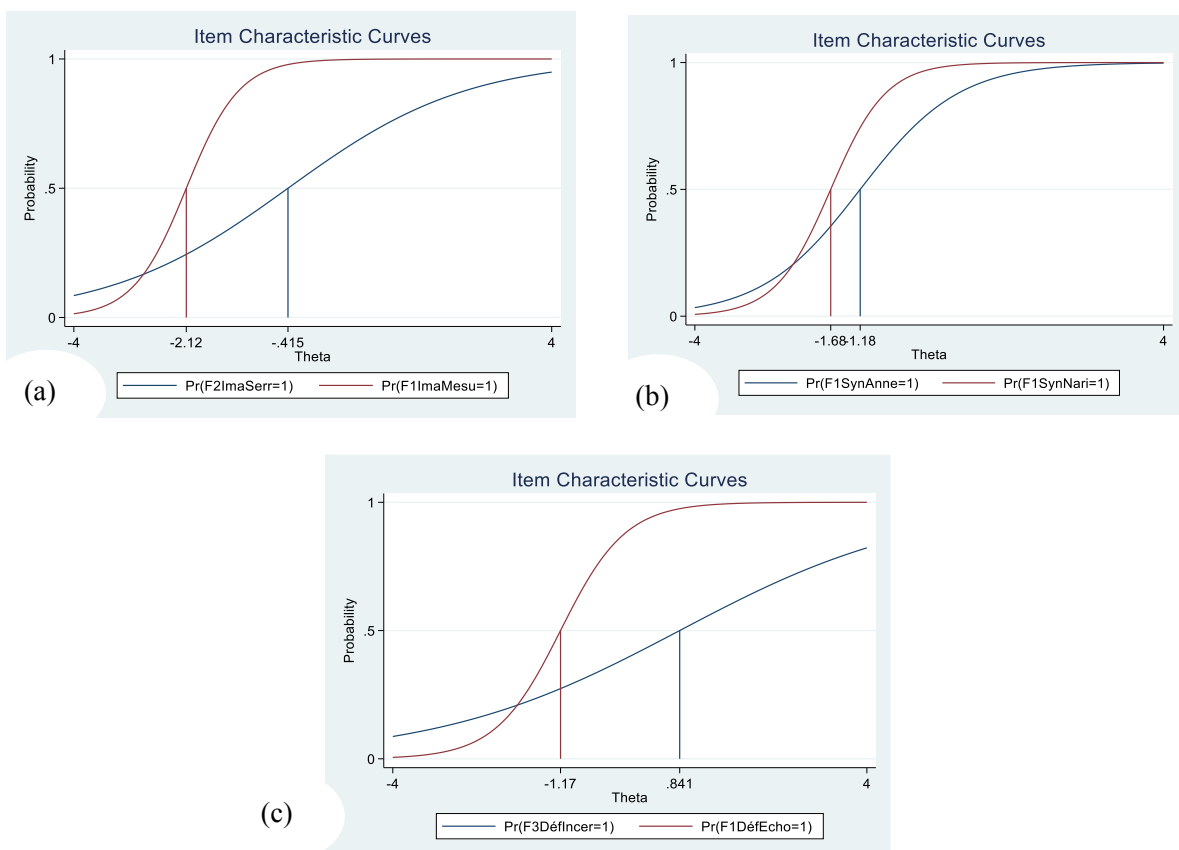


Tableau 7*Statistiques des indices d'ajustement des modèles M1 et M2 des analyses factorielles confirmatoires*

| | χ^2 | df | RMSEA [90 % CI] | CFI | TLI | SRMR | AIC | BIC | $\Delta\chi^2$ | \neq Modèle |
|----|----------|----|---------------------|-----|------|------|------|------|-----------------|---------------|
| M1 | 4,16 | 4 | ,013 [,000 ,096] | 1 | ,999 | ,013 | 3146 | 3203 | - | - |
| M2 | 3,95 | 5 | ,000 [,000 ,077] | 1 | 1 | ,013 | 3144 | 3197 | $\chi^2(1)=,21$ | 1 vs 2 |

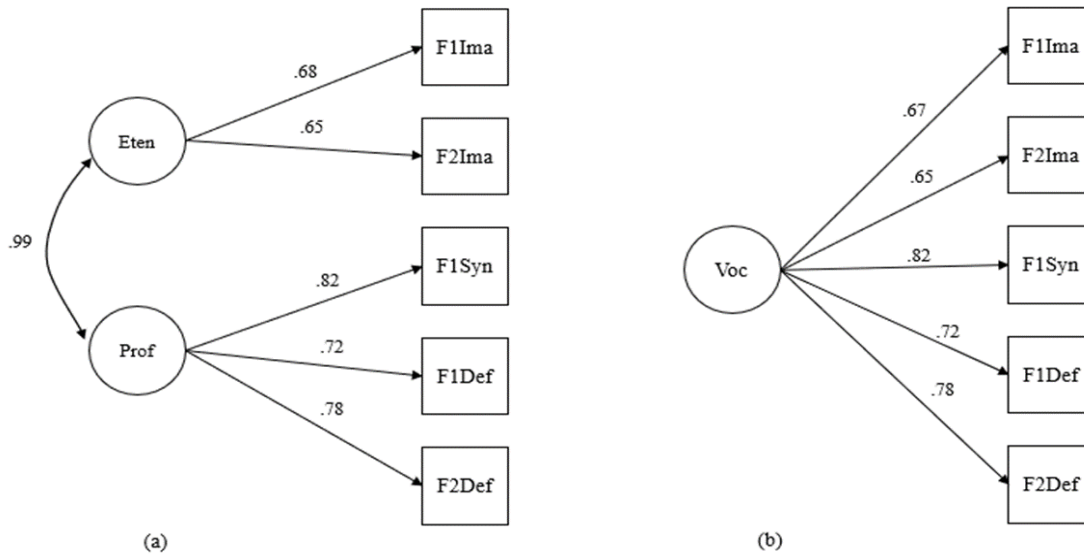
Note : $\Delta\chi^2$ = chi square Satorra-Bentler.

Afin de déterminer la meilleure structure factorielle à partir des scores dans les trois tâches et en fonction de la fréquence lexicale (soit les 36 items répartis en cinq catégories de scores⁵; voir Figure 4), nous avons comparé notre modèle initial M1 à deux facteurs au modèle alternatif M2 à un seul facteur. Nous nous attendons à ce que M1 structuré en deux facteurs latents correspondant à l'étendue et la profondeur du vocabulaire ait un meilleur ajustement aux données que M2 avec un seul facteur intégrant l'ensemble des scores du test, c'est-à-dire représentant le vocabulaire. Les deux modèles M1 et M2 révèlent des indices d'ajustements excellents (Tableau 7). Une analyse comparative entre M1 et M2 n'indique pas de différence significative ($\Delta\chi^2 = ,21, p > ,10$) en faveur de M1 conformément à notre hypothèse. Ces résultats ne permettent pas de retenir ou de rejeter l'une des structures factorielles entre M1 (bidimensionnelle) et M2 (unidimensionnelle) (Figure 4).

⁵ Les items retenus dans la tâche Synonyme sont tous de fréquence élevée (F1).

Figure 4

Coefficients standardisés des analyses factorielles confirmatoires du Modèle 1 en deux facteurs (a) et du Modèle 2 avec un facteur (b)



Notes : Eten : Étendue; Prof : Profondeur, Voc : Vocabulaire; Ima : Image; Syn : Synonyme; Def : Définition; F1 : mots à fréquence élevée; F2 : mots à fréquence faible.

Discussion

L'objectif de cette étude était de concevoir et développer un test de vocabulaire utilisant la technologie numérique pour des enfants de la deuxième année de primaire (G2) à la cinquième année (G5). Le dispositif novateur qui a été mis en œuvre devait faire face à plusieurs défis : 1. tester les enfants au sein des écoles primaires dans les salles équipées d'ordinateurs connectés à Internet puisque le logiciel d'évaluation était sur un serveur, 2. utiliser plusieurs tâches pour évaluer les deux dimensions du vocabulaire, et 3. sélectionner un nombre restreint de mots pour ne pas rendre la tâche trop longue et à la fois sélectionner des mots en fonction de leur catégorie grammaticale (noms, verbes et adjectifs) et de leur fréquence lexicale.

Effet des caractéristiques individuelles et lexicales

Globalement, les variables individuelles (niveau scolaire et genre) et lexicales (fréquence des mots) convoquées, impactant les performances (vitesse et/ou exactitude), ont confirmé les hypothèses avancées. Tout d'abord, les effets de l'âge (ici en ce qui a trait au niveau scolaire) et de la fréquence lexicale ont été mis en évidence sur les réponses correctes dans les trois tâches (« Image », « Définition », « Synonymes »). Le niveau de vocabulaire augmente avec l'âge (p. ex., Rice & Hoffman, 2015). Les Rc étaient plus élevées pour les mots fréquents. De plus, l'effet du genre est, quant à lui, moins marqué; les filles ayant significativement des performances supérieures à celles des garçons seulement dans la tâche « Définition ». À l'instar de Fenson *et al.* (2007), les filles

semblent avoir un vocabulaire plus riche que celui des garçons. Ces résultats suggèrent de tenir compte du genre de façon systématique lorsqu'on évalue le niveau de vocabulaire, ou encore afin par exemple d'adapter les normes de tests. Par ailleurs, les analyses sur les Rc révèlent un effet d'interaction significatif entre l'âge et la fréquence dans la tâche « Synonyme », où comme attendu, les Rc augmentent avec l'âge et avec la fréquence lexicale. Il est plus facile de répondre correctement pour les mots fréquents comparativement aux mots rares, et ce d'autant plus lorsque les enfants ont un niveau scolaire plus élevé. Dans le cadre de l'hypothèse de la qualité lexicale, les représentations lexicales seraient suffisamment stables et précises pour récupérer le sens des mots et sélectionner le synonyme (p. ex., Perfetti & Stafura, 2014).

Les analyses portant sur les TR montrent des effets moins marqués des temps sur les différentes variables. En effet, on observe bien l'effet attendu du niveau scolaire (confondu avec l'âge) sur les temps, les temps décroissant de G2 à G5, mais uniquement pour l'épreuve Image. Pour le Genre, on observe un effet des TR seulement dans la tâche Définition en faveur des garçons qui répondent plus vite. Enfin, l'effet de Fréquence sur les temps est observé dans l'épreuves Image en faveur des mots plus fréquents traités plus vite ce qui est l'inverse dans la tâche Synonyme où ce sont les mots moins fréquents qui sont traités plus rapidement. Pour ce dernier résultat inattendu, il faudrait également prendre en compte la fréquence des items tests proposés. En effet, c'est bien la fréquence des items sources (items pour lesquels il faut trouver le synonyme dans la liste; p. ex. : écorce) qui a été prise en compte mais pas la fréquence des items tests et notamment celle de l'item cible (p. ex. : peau). Enfin, les analyses ont révélé un effet d'interaction significatif entre le genre et la fréquence lexicale sur les TR dans la tâche « Image ». L'effet de fréquence lexicale était plus important pour les garçons plutôt que pour les filles.

Qualités psychométriques

Construire un outil dont l'objectif est de devenir un test impose d'en examiner les qualités psychométriques (Laveault & Grégoire, 2014). Leur examen, réalisé en trois temps (rejet des items aux réponses correctes ne se distinguant pas du hasard; calcul des r_{pbis} ; calcul des coefficients de discrimination avec le MRI), a conduit à réduire l'épreuve à 36 items (sur les 85 items initiaux). Cette sélection certes sévère permet alors d'obtenir un coefficient élevé de consistance interne de l'épreuve. Toutefois, nous pouvons nous demander si une épreuve de vocabulaire réduite à 36 items peut évaluer de façon complète et satisfaisante le vocabulaire. Il semble évident que d'autres items avec les mêmes contrôles (catégorie grammaticale, fréquence lexicale) soient nécessaires pour compléter l'épreuve. Il faudrait alors procéder à de nouvelles analyses portant sur la qualité des items. On voit toute la complexité et le temps long nécessaire pour construire une telle épreuve au contenu satisfaisant.

Un tri également strict a aussi porté sur les TR. C'est un aspect important dans le contexte numérique puisque l'objectif d'une telle épreuve est d'obtenir les paramètres vitesse et précision pour récupérer le sens des mots (p. ex., Richter *et al.*, 2013). Ces deux paramètres sont importants à prendre en compte (p. ex., Perfetti, 2007) pour estimer le poids du vocabulaire en lecture, sur les deux composantes, identification de mots écrits et compréhension (voir Oakhill *et al.*, 2019 pour une revue). Des données sur le vocabulaire intégrant ces deux paramètres peuvent encore mieux expliquer les performances en lecture. Des travaux l'ont montré, et d'autres sont en cours sur ce point. Richter *et al.* (2013) ont montré que la précision de la signification des mots influençait

indirectement la compréhension en lecture, en raison de la précision des représentations orthographiques et phonologiques des mots. Ainsi, la précision renvoie à l'exactitude avec laquelle les mots, mais aussi leur signification, sont stockés, et la vitesse renvoie à la rapidité avec laquelle les mots et leur signification sont récupérés. La combinaison de ces deux paramètres permet alors de rendre compte de l'efficacité des processus impliqués (p. ex., Perfetti & Stafura, 2014).

Analyses des deux dimensions du vocabulaire

Pour finir, lors de la construction du test, on faisait l'hypothèse que la tâche « Image » relevait de la dimension « étendue » alors que les tâches « Définition » et « Synonyme » devraient plutôt renvoyer à la profondeur du vocabulaire (voir Pearson *et al.*, 2007). Nous faisons donc l'hypothèse que les données de cette épreuve permettraient d'extraire ces deux dimensions. Toutefois, les analyses factorielles confirmatoires n'ont pas permis de confirmer cette hypothèse. Sans remettre en question cette conceptualisation du vocabulaire, il semblerait qu'elle soit plus difficile à mettre en évidence. Toutefois, la profondeur du vocabulaire qui renvoie aux multiples connaissances associées à un mot mérite d'être investiguée. D'autres travaux sur la création d'épreuves appropriées et avec de nouveaux items sont à développer pour évaluer de façon plus complète les deux composantes du vocabulaire.

Conclusion

La construction d'un outil informatisé pour évaluer une compétence dans le domaine des apprentissages nécessite du temps et de nombreuses analyses avant d'obtenir un outil solide et fiable. Cette étude au caractère exploratoire avait pour objectif de montrer les différentes phases qui nous ont conduits de l'examen des caractéristiques individuelles et linguistiques sur les performances en vocabulaire aux qualités psychométriques dont il nous semble essentiel d'en fournir les éléments. De ces dernières analyses, essentielles, dépendent in fine la sélection finale des items. À l'issue de ce premier travail proposé sur ordinateur, une question subsiste : faut-il dès maintenant développer ce type de tests sur tablette à la gestion motrice et cognitive plus intuitive, donc plus simple? Poser ainsi cette question nous semble d'ores et déjà y répondre : des tentatives récentes ont vu déjà le jour dans le domaine (voir par ex., Dujardin *et al.*, 2021; Schaefer *et al.*, 2016, 2019).

Ainsi, évaluer le vocabulaire avec un dispositif numérique nous semble très prometteur. En effet, cette dimension nécessite en particulier dans le domaine de l'apprentissage de la lecture d'être prise en compte au-delà des composantes d'identification de mots écrits et de compréhension (Magnan & Écalle, 2018). En d'autres termes, en élargissant l'examen des causes susceptibles d'être à l'origine des difficultés dans l'apprentissage et la maîtrise de la lecture, ce type de travaux de recherche se proposent de fournir à moyen terme aux praticiens (enseignants, rééducateurs, orthophonistes) un nouvel outil numérique centré sur le vocabulaire dont on connaît l'importance en lecture et dans tous les apprentissages.

Remerciements

Ce travail a été conduit dans le cadre d'une thèse CIFRE (Conventions industrielles de formation par la recherche) réalisée par le second auteur en partenariat avec le Réseau Canopé Lyon

sous la direction du premier et du dernier auteur. Nous remercions Samuel Bocharel, responsable éditorial à Canopé et François Morel, directeur territorial adjoint Réseau Canopé Auvergne-Rhône-Alpes (AuRA) pour leur soutien. Nos remerciements vont également au Professeur Roger Fougères, chargé de mission pour la mise en place de l'Institut Carnot Education (AuRA) pour son engagement dans ce projet de thèse

Références

- Auphan, P., Écalle, J., & Magnan, A. (2019). Subtypes of readers with a computer-based assessment of reading ability: A study in French children from G2 to G9. *European Journal of Psychology of Education*, 34(3), 641-663. <https://doi.org/10.1007/s10212-018-0396-7>
- Auphan, P., Écalle, J., & Magnan, A. (2020). The high potential of computer-based reading assessment. *Canadian Journal of Learning and Technology*, 46(1), 1-23. <https://doi.org/10.21432/cjlt27847>
- Beauvais, L., Bouchafa, H., Beauvais, C., Kleinsz, N., Magnan, A., & Écalle, J. (2018). Tinfolec : A new French web-based test for reading assessment in primary school. *Canadian Journal of School Psychology*, 33(3), 227-241. <https://doi.org/10.1177/0829573518771130>
- Burnham, K. P., Anderson, D. R., & Burnham, K. P. (2002). *Model selection and multimodel inference : A practical information-theoretic approach* (2e éd.). Cham : Springer.
- Cunningham, A. E. (2005). Vocabulary growth through independent reading and reading aloud to children. Dans E. H. Hiebert et M. L. Kamil, *Teaching and learning vocabulary. Bringing research to practice* (p. 45- 68). Mahwah, NJ : Lawrence Erlbaum Associates Publishers.
- Dujardin, E., Écalle, J., Auphan, P., Gomes, C., Cros, L., & Magnan, A. (2021). Vocabulary assessment with tablets in Grade 1 : Examining effects of individual and contextual factors and psychometric qualities. *Frontiers in Education*, 6:664131. <https://doi.org/10.3389/educ.2021.664131>
- Dunn A., & Dunn A. (1981). *Peabody picture vocabulary test-revised*. Circle Pines, MN : American Guidance Service.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 : Peabody picture vocabulary test*. San Antonio, TX : Pearson Assessments.
- Dunn, L. M., Dunn, D. M., & Styles, B. (2009). *The british picture vocabulary scale*. Brentford : GL Assessment Limited.
- Dunn, L. M., Dunn, L. M., & Theriault-Whalen, C. (1993). *EVIP : Échelle de vocabulaire en images Peabody [Pictures vocabulary scale Peabody]*. Toronto, ON : Pearson Assessments.
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates communicative development inventories*. Baltimore, MD : Brookes.
- Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development*, 17(1), 1- 17. <https://doi.org/10.1080/15248372.2015.1061528>
- Good, R. H., & Kaminski, R. A. (2002). *Dynamic indicators of basic early literacy skills : DIBELS*. Dans R. H. Good (dir.), *Dynamic measurement group*. Eugene, OR : DIBELS.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26(1), 55- 88. <https://doi.org/10.1016/j.dr.2005.11.002>

- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis : Conventional criteria versus new alternatives. *Structural Equation Modeling. A Multidisciplinary Journal*, 6(1), 1- 55. <https://doi.org/10.1080/10705519909540118>
- Laveault, D., & Grégoire, J. (2014). *Introduction aux théories des tests en sciences humaines* (3e éd.). Bruxelles : De Boeck.
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX : A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers*, 36(1), 156-166. <https://doi.org/10.3758/BF03195560>
- Magnan, A., & Écalte, J. (2018). L'apprentissage de la lecture : des processus d'identification de mots écrits à la compréhension. Dans R. Baldy (dir.), *Dessiner, lire, écrire et calculer : un regard neuf* (p. 93-172). Paris : In Press
- Marble-Flint, K. J., Strattman, K. H., & Schommer-Aikins, M. A. (2019). Comparing iPad ® and paper assessments for children with ASD : An initial study. *Communication Disorders Quarterly*, 40(3), 152- 155. <https://doi.org/10.1177/1525740118780750>
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus : Statistical analysis with latent variables. User's Guide* (Version 8). Los Angeles, CA : Authors.
- Neumann, M. M., & Neumann, D. L. (2019). Validation of a touch screen tablet assessment of early literacy skills and a comparison with a traditional paper-based assessment. *International Journal of Research & Method in Education*, 42(4), 385- 398. <https://doi.org/10.1080/1743727X.2018.1498078>
- Oakhill, J., Cain, K., & Elbro, C. (2019). Reading comprehension and reading comprehension difficulties. Dans D. A. Kilpatrick, R. M. Joshi et R. K. Wagner (dir.), *Reading development and difficulties. Bridging the gap between research and practice* (p. 83- 115). Cham : Springer International Publishing. https://doi.org/10.1007/978-3-030-26550-2_5
- Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment : What we know and what we need to learn. *Reading Research Quarterly*, 42(2), 282- 296. <https://doi.org/10.1598/RRQ.42.2.4>
- Perfetti, C. (2007). Reading ability : Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357- 383. <https://doi.org/10.1080/10888430701530730>
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22- 37. <https://doi.org/10.1080/10888438.2013.827687>
- Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment : A longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language, and Hearing Research*, 58(2), 345- 359. https://doi.org/10.1044/2015_JSLHR-L-14-0150
- Richter, T., Isberner, M.-B., Naumann, J., & Neeb, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading*, 17(6), 415- 434. <https://doi.org/10.1080/10888438.2013.764879>

- Schaefer, B., Bowyer-Crane, C., Herrmann, F., & Fricke, S. (2016). Development of a tablet application for the screening of receptive vocabulary skills in multilingual children : A pilot study. *Child Language Teaching and Therapy*, 32(2), 179- 191. <https://doi.org/10.1177/0265659015591634>
- Schaefer, B., Ehlert, H., Kemp, L., Hoesl, K., Schrader, V., Warnecke, C., & Herrmann, F. (2019). Stern, gwiazda or star : Screening receptive vocabulary skills across languages in monolingual and bilingual German–Polish or German–Turkish children using a tablet application. *Child Language Teaching and Therapy*, 35(1), 25- 38. <https://doi.org/10.1177/0265659018810334>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*, 99(6), 323- 338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Sénéchal, M., Whissell, J., & Bildfell, A. (2017). Starting from home: Home literacy practices that make a difference. Dans K. Cain, D. L. Compton et R. K. Parrila (dir.), *Studies in written language and literacy* (vol. 15, p. 383- 408). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/swll.15.22sen>
- Simos, P. G., Sideridis, G. D., Protopapas, A., & Mouzaki, A. (2011). Psychometric evaluation of a receptive vocabulary test for greek elementary students. *Assessment for Effective Intervention*, 37(1), 34- 49. <https://doi.org/10.1177/1534508411413254>
- Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.) Boston, MA: Pearson Education.
- Taylor, C. L., Christensen, D., Lawrence, D., Mitrou, F., & Zubrick, S. R. (2013). Risk factors for children’s receptive vocabulary development from four to eight years in the longitudinal study of Australian children. *PLoS ONE*, 8(9), e73046. <https://doi.org/10.1371/journal.pone.0073046>
- Wiig, E., Secord, W., Semel, E., Boulianne, L., & Labelle, M. (2009). *Évaluation clinique des notions langagières fondamentales : Version pour francophones du Canada* (Clinical evaluation of language fundamentals: French Canadian version). Toronto, ON: Pearson Assessments.
- Wiig, E. H., & Secord, W. (1992). *TOWK. Test of word knowledge*. Toronto, ON : Psychological Corporation.
- Wiig, E. H., Secord, W. A., & Semel, E. (2013). *Clinical evaluation of language fundamentals: CELF-5*. Toronto, ON: Pearson Assessments.
- Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research*, 50, 1562-1592. [https://doi.org/10.1044/1092-4388\(2007/106\)](https://doi.org/10.1044/1092-4388(2007/106))

Auteurs

Jean Ecalle est professeur émérite au sein du Laboratoire EMC de l'Université Lyon 2. Ses travaux portent sur l'apprentissage de la lecture auprès de populations d'enfants "tout-venant" et d'enfants en difficultés d'apprentissage. Une partie de ses recherches porte spécifiquement sur l'évaluation des performances en lecture et l'utilisation d'aide informatisée en lecture (voir www.ecalle-magnan.fr). Courriel : ecalle.jean@wanadoo.fr

Nicolas Bailloud est docteur en Sciences Cognitives. Il travaille en tant que médiateur Formation à l'atelier Canopé de Lyon, et a un fort intérêt pour les neurosciences de l'éducation. Ses travaux portent, notamment, sur le développement d'applications Web pour évaluer et stimuler le vocabulaire chez l'enfant. Courriel : nicolas.bailloud@reseau-canope.fr

Emilie Dujardin est docteure en Psychologie Cognitive et du Développement. Elle est en contrat postdoctoral, financé par l'Association Agir pour l'Ecole, au sein du laboratoire EMC. Ses travaux portent sur l'apprentissage de la lecture de l'enfance à l'âge adulte, et visent à déterminer les différences inter-individuelles. Courriel : emilie.dujardin@univ-lyon2.fr

Annie Magnan est professeure émérite au sein du Laboratoire EMC de l'Université Lyon 2, membre senior honoraire de l'Institut Universitaire de France. Ses travaux portent sur l'apprentissage de la lecture chez l'enfant. Ses recherches visent à évaluer les performances en lecture, aider à l'apprentissage de la lecture via des remédiations (ex. : aide informatisée) auprès d'enfants au développement normal et d'enfants présentant des difficultés d'apprentissage de la lecture (dyslexiques, sourds, porteurs du Syndrome de Down...). Pour accéder à l'ensemble des publications (voir www.ecalle-magnan.fr). Courriel : annie.magnan@univ-lyon2.fr



This work is licensed under a Creative Commons Attribution-NonCommercial CC-BY-NC 4.0 International license.