DEPARTEMENT ORTHOPHONIE
FACULTE DE MEDECINE
Pôle Formation
59045 LILLE CEDEX
Tél : 03 20 62 76 18
departement-orthophonie@univ-lille.fr

MEMOIRE

En vue de l'obtention du

Certificat de Capacité d'Orthophoniste

présenté par

Jenna Delmas

soutenu publiquement en juin 2024

Le bilan de cognition mathématique :
Validation d'une épreuve de comparaison de nombres arabes

MEMOIRE dirigé par

Sandrine MEJIAS, Maître de conférences, Département d'orthophonie, Université de Lille, SCALab

Sophie FRAGNON, Orthophoniste, Wingles et enseignante, Département d'orthophonie, Université de Lille

Lille – **2024**

Remerciements

Je tiens particulièrement à remercier Madame Mejias et Madame Fragnon, mes directrices de mémoire, pour m’avoir fait confiance dans la poursuite de ce projet ainsi que pour leur disponibilité et leurs conseils précieux et avisés.

Mes remerciements vont également aux nombreuses personnes qui ont accepté de participer à ce projet. Je remercie notamment les équipes enseignantes pour leur accueil et le temps qu’elles m’ont accordé, les élèves pour leur motivation et leur patience, ainsi que leurs parents pour l’intérêt qu’ils ont porté à cette étude. Ce mémoire n’aurait pas pu aboutir sans leur aide.

Pour l’expérience clinique apportée, pour la confiance accordée, mais surtout pour leur profonde bienveillance, je tiens à exprimer ma gratitude envers mes maîtres de stage de troisième, de quatrième et de cinquième année.

Un immense merci à Charlotte, Lucie, Chloé et Angèle, sans qui ces cinq années d’études n’auraient pas été aussi belles.

J’adresse mes plus profonds remerciements à mes parents, ma sœur et mon frère, ainsi qu’à tout le reste de ma famille. Leur patience, leur écoute, leur réconfort et leur soutien affectif ont été et sont toujours d’une valeur inestimable. Merci de toujours croire en moi.

Enfin, un sincère merci à mes ami.e.s pour leur présence depuis toutes ces années, leur soutien inconditionnel et leurs encouragements au quotidien.

Résumé :

Considérée comme une habileté numérique de base, la capacité à comparer deux nombres arabes est primordiale pour un développement mathématique abouti et prédit les capacités arithmétiques futures. En ce sens, investiguer les compétences en comparaison symbolique est précieux lors d'une évaluation des habiletés mathématiques d'un enfant. Dans un contexte de projet de création d'une nouvelle batterie d'évaluation de la cognition mathématique, notre étude concerne la phase de pré-test d'une épreuve de comparaison de nombres arabes. Ce travail initial de validation avait pour objectif d'évaluer la pertinence et la qualité de cette épreuve, à travers des données qualitatives et quantitatives préliminaires. Au total, 79 sujets témoins scolarisés du CP au CM2 ont passé ce test. Les résultats révèlent la rapidité d'application et la clarté des instructions, et suggèrent que les items sont conformes aux performances attendues selon le niveau des enfants. Cependant, une poursuite de ce travail est nécessaire. Elle devra inclure un échantillon plus important afin d'aboutir à une décision finale concernant les ajustements nécessaires en termes de présentation, de contenu et de cotation de l'épreuve.

Mots-clés :

Cognition mathématique, évaluation, pré-test, comparaison symbolique.

Abstract :

Considered a basic numerical skill, the ability to compare two arabic numbers is essential for successful mathematical development and predicts future arithmetic abilities. As such, investigating symbolic sombolic comparison skills is invaluable when assessing a child's mathematical abilities. In the context of a project to create a new battery for assessing mathematical cognition, our study concerns the pre-test phase of an arabic number comparison test. The aim of this initial validation work was to assess the relevance and quality of this test, through preliminary qualitative and quantitative data. A total of 79 control subjects from CP to CM2 took the test. The results show that the test is applied quickly and the instructions are clear, and suggest that the items are in line with expected performance at the children's level. However, this work needs to be continued. It will need to include a larger sample in order to reach a final decision on the adjustments needed in terms of test presentation, content and scoring.

Keywords :

Mathematical cognition, assessment, pre-test, symbolic comparison.

Table des matières

Introduction

La comparaison de nombres arabes est une compétence mathématique fondamentale, présente dans de nombreuses activités de la vie quotidienne, et nécessaire à la prise de décisions éclairées. L'acquisition de cette habileté se produit dès le plus jeune âge et relève d'un processus complexe, dynamique et continu, impliquant à la fois des habiletés innées et des compétences acquises (Mundy & Gilmore, 2009). Enseignée dès la maternelle, la capacité à comparer deux codes symboliques arabes est considérée comme une habileté numérique de base, essentielle au développement des compétences mathématiques futures, telles que le calcul, la résolution de problèmes, la géométrie… (De Smedt et al., 2009; Hawes et al., 2019).

Le Trouble des Apprentissages Mathématiques (TAM) est un trouble spécifique des apprentissages, dont la prévalence est estimée entre 2% et 8% selon les critères diagnostiques et les tests utilisés (Wahl & Wahl, 2020). Malgré ce taux non négligeable, peu d'études font l'objet de recherches sur les déficits en lien avec ce trouble, et leur nature. Des études ont permis d'identifier des difficultés spécifiques en comparaison de quantités symboliques chez les enfants porteurs d'un TAM, suggérant que les résultats obtenus à cette tâche pourraient constituer un indicateur central pour diagnostiquer un TAM de façon précoce (Schwenk et al., 2017). Une évaluation des habiletés numériques de base, notamment de la comparaison symbolique, apparaît donc indispensable afin d'établir un diagnostic de TAM. L'orthophoniste possède les qualifications pour évaluer ces aptitudes. Pourtant, évaluer précisément les déficits numériques reste un défi, car les outils dont il dispose ne répondent pas tous à de bonnes qualités théoriques et psychométriques (Lafay & Cattini, 2018). Afin de garantir une rééducation conforme au profil cognitif du patient, un diagnostic précis et basé sur des preuves solides est essentiel pour identifier les déficits numériques qui lui sont propres.

Ce mémoire s'inscrit dans un projet de création d'une batterie d'évaluation diagnostique de la cognition mathématique, fondée sur des modèles théoriques actuels. Par son travail antérieur, Cousin (2021) a proposé la création d'une épreuve de comparaison de nombres arabes. Innovante dans son contenu, elle se veut également exhaustive grâce à la présence de soixante items. Une telle épreuve permettrait à l'orthophoniste d'objectiver d'éventuelles difficultés en lien avec l'efficacité du système numérique approximatif, la précision numérique et le développement du sens du nombre. Actuellement, nous participons à la validation de cette épreuve, en passant par l'analyse et l'interprétation des résultats obtenus lors de premières passations. Cette phase de pré-test permettra de vérifier la pertinence et la qualité de ce test, et d'identifier d'éventuels ajustements.

Dans un premier temps, nous exposerons le cadre théorique à partir duquel nous avons construit notre travail. Après avoir présenté le modèle du triple code de Dehaene (1992), nous nous intéresserons plus spécifiquement à la comparaison de nombres arabes, en évoquant les aspects développementaux et cognitifs de cette compétence. Nous finirons par préciser son lien avec le TAM, puis justifier l'intérêt de son évaluation. Il s'agira ensuite d'exposer l'objectif de cette étude, ses hypothèses et sa méthodologie. Enfin, nous détaillerons les résultats que nous discuterons par la suite.

Contexte théorique, buts et hypothèses

Dans cette partie, nous rappellerons l'importance du modèle du triple code de Dehaene (1992) dans la compréhension des fondements de la cognition numérique. Nous nous appuierons ensuite sur la littérature actuelle, pour définir la comparaison de nombres arabes et décrire les aspects développementaux et cognitifs de cette compétence. Puis, il s'agira de s'intéresser au TAM en lien avec la comparaison de nombres arabes, et de préciser l'intérêt d'une telle épreuve dans la démarche diagnostique. Enfin, nous formulerons les objectifs et hypothèses de ce mémoire.

1. L'apport du triple code de Dehaene dans la cognition numérique

Le modèle du triple code (Dehaene, 1992) fournit un cadre théorique pour comprendre la manière dont les humains se représentent et manipulent les quantités numériques.

1.1. Le modèle et les représentations numériques associées

Le modèle anatomique et fonctionnel de Dehaene présume l'existence de trois systèmes de représentation mentale des nombres chez l'adulte : le code analogique, le code auditivo-verbal et le code visuel-arabe (Dehaene et al., 2003).

Le code analogique réfère à la représentation non symbolique des nombres et soutient la capacité innée de traiter de manière approximative les quantités. Impliqué dans la comparaison des quantités et les calculs approximatifs, ce système permet une évaluation précise des petites quantités et une estimation approximative des grandes collections. La représentation analogique serait alors porteuse du sens du nombre (Dehaene, 2010). Elle active le cortex pariétal inférieur gauche, plus précisément le sulcus intrapariétal (Dehaene et al., 1999).

Les codes auditivo-verbal et visuel-arabe sont quant à eux, des systèmes de représentation symbolique qui permettent respectivement de se représenter les quantités à travers les nombres verbaux ou à travers la forme visuelle des nombres arabes. Le code verbal intervient lors d'une tâche de comptage ou une tâche impliquant les faits arithmétiques. Il active les mêmes aires qui sont impliquées dans le traitement du langage, dont le gyrus angulaire gauche (Dehaene et al., 1999). Le code arabe est quant à lui important pour la lecture et l'écriture des nombres, les activités de calculs mentaux complexes ou les jugements de parité. Il est impliqué dans la reconnaissance des nombres et la comparaison de leur magnitude. Il active quant à lui, les aires occipito-temporales bilatérales (Dehaene et al., 1999). Les deux systèmes symboliques permettent donc de réaliser des calculs précis au-delà des petites quantités.

Ce modèle permet d'expliquer que les trois représentations sont à la fois indépendantes et en interaction constante. Cependant, il n'apporte pas de précision concernant la mise en place des représentations symboliques et l'établissement des liens entre les codes.

1.2. Les représentations numériques dans une perspective développementale

Dès l'âge de 6 mois, le nourrisson est capable d'estimer, discriminer et manipuler des quantités de manière approximative (Lipton & Spelke, 2003; Xu & Spelke, 2000). Il peut distinguer un petit nombre d'objets d'un grand nombre (par exemple une collection de trois pommes d'une collection de six pommes). En d'autres termes, un nourrisson sait faire la distinction entre " un peu" et " beaucoup". Les enfants commenceraient donc par utiliser des représentations analogiques pour évaluer la taille et la distance entre les quantités numériques. Cette représentation analogique des quantités serait ainsi innée, et rendue possible par deux systèmes distincts : le Système Numérique Précis (SNP) responsable du traitement rapide des petites quantités, et le Système Numérique Approximatif (SNA) responsable du traitement approximatif des grandes quantités (Feigenson et al, 2004).

Grâce au langage et à la confrontation aux annotations symboliques, les enfants accèdent ensuite aux codes symboliques. Cette représentation symbolique est spécifique aux humains et dépendante d'un apprentissage plus ou moins explicite. Lorsque cet apprentissage est acquis, les codes symboliques s'intègrent à la représentation analogique et commencent à leur tour à véhiculer du sens (Chazoule & Thevenot, 2018; Lipton & Spelke, 2005; Mundy & Gilmore, 2009). C'est donc grâce aux systèmes verbal et arabe que vers l'âge de 2-3 ans, les enfants apprennent à nommer les nombres et à se représenter de façon précise des quantités précises à partir des symboles (Rousselle et al., 2004).

Les représentations numériques évoluent ensuite au cours du développement, pour se préciser à l'âge adulte (Piazza et al., 2010).

2. Focus sur la comparaison de nombres arabes

L'acquisition de la comparaison symbolique arabe est une étape importante dans le développement de la cognition mathématique. La comparaison de nombres arabes désigne la faculté à déterminer si un nombre est plus grand, plus petit ou égal à un autre nombre, en utilisant uniquement la représentation arabe écrite. Autrement dit, il s'agit de la faculté à comprendre la relation numérique entre deux nombres arabes.

2.1. Aspects développementaux

2.1.1. Le développement de la capacité à comparer deux nombres arabes

Comme mentionné précédemment, le développement des codes s'effectue de manière progressive. Cela implique le passage par plusieurs étapes de développement avant d'acquérir la faculté à comparer deux nombres arabes.

Le traitement des nombres arabes s'appuierait sur le système analogique déjà existant. Cela est notamment soutenu par l'étude de Piazza et al. (2007) qui a permis de démontrer que la région cérébrale liée au sens du nombre, le sillon intrapariétal, était activée lors de tâches faisant intervenir le code arabe. L'acquisition de la faculté à comparer des nombres arabes prendrait alors appui sur des étapes précédentes de compréhension des quantités. Les expérimentations menées par Chazoule et Thevenot (2018) vont également dans ce sens. Ces auteurs ont mis en évidence une amélioration progressive de la capacité à effectuer des comparaisons numériques en fonction de l'âge. À l'âge de 3 ans, les enfants ne maîtrisent pas encore les codes symboliques ; leurs performances sont meilleures en comparaison analogique de jetons. Entre l’âge de 3 ans et 6 ans, l'acquisition des codes symboliques verbaux et arabes est particulièrement remarquable : les performances en comparaison de numéraux verbaux rejoignent, puis dépassent les performances aux épreuves analogiques à l’âge de 5 ans. Les capacités en comparaison symbolique arabe continuent quant à elles de s’accroitre avec l’âge, et deviennent équivalentes aux capacités de comparaison analogique et symbolique verbale à l’âge de 6 ans.

Les enfants commencent donc à être capables de comparer des nombres via le code numérique arabe à partir de l’âge de 5-6 ans (White et al., 2012). Au cours des acquisitions, ils développent une meilleure compréhension de la magnitude numérique. Ils deviennent capables de comparer de manière plus rapide et plus précise des nombres de plus en plus grands, mais également des nombres plus complexes comme les fractions et décimaux. Cette capacité émerge plus tardivement et dépend de la compréhension de la position des chiffres dans le nombre. Une étude a montré que des difficultés à comprendre les fractions pouvaient persister jusqu’au lycée (Siegler & Pyke, 2013).

La progression dépend de plusieurs facteurs et ne suit pas forcément un développement linéaire. Toutefois, la comparaison de nombres arabes se développe dès le plus jeune âge et fait intervenir à la fois des représentations analogiques et symboliques et la cartographie entre les deux systèmes (Mundy & Gilmore, 2009). En effet, il est évident qu’une tâche de comparaison de nombres arabes repose sur le code indo-arabe. Celui-ci permet d’extraire des informations visuelles, de reconnaître et de comprendre les symboles. Le code analogique, lui, s’activerait afin de représenter et estimer la magnitude de chaque nombre. Il serait alors possible de déterminer quelle quantité est la plus grande grâce à la comparaison des deux magnitudes numériques.

2.1.2. La place de la comparaison de nombres arabes dans les programmes scolaires

Les attendus scolaires en mathématiques, issus du bulletin n°31 relatif aux programmes d’enseignements primaire et secondaire (Ministère de l’Education Nationale et de la Jeunesse, 2020), sont en accord avec les données de la littérature.

Dès le cycle 1, l’évaluation et la comparaison de collections sont enseignées. Les enfants doivent être capables de comparer des quantités inférieures ou égales à 10 à partir du dénombrement. À la fin de l’école maternelle, ils commencent à savoir comparer deux nombres arabes inférieurs ou égaux à 10. Au cours du cycle 2, il est stipulé que les élèves doivent être capables de dénombrer, ordonner dans l’ordre croissant et décroissant et comparer, à partir de nombres entiers. Ces compétences se développent à partir de plusieurs supports tels que l’utilisation de symboles ou d’expressions verbales de comparaison ("égal à", "plus que", "plus grand que"...). Le cours préparatoire (CP) est une année déterminante où les compétences numériques de base sont enseignées et où les trois systèmes de représentation du nombre doivent être mis en relation. Progressivement et de manière respective, les élèves de cours préparatoire (CP), cours élémentaire première année (CE1) et cours élémentaire deuxième année (CE2) doivent étendre ces compétences à des nombres entiers inférieurs ou égaux à 100, 1 000, puis 10 000. Le cycle 3, dans la continuité du cycle 2, approfondit le développement de ces compétences majeures en introduisant de nouvelles notions, plus complexes. Ainsi, les élèves de cours moyen première année (CM1), cours moyen deuxième année (CM2) et 6ème doivent non seulement utiliser et représenter des nombres entiers plus grands, mais aussi des fractions simples et des nombres décimaux. En d’autres termes, ils doivent être capables de les positionner sur une ligne numérique, de les comparer et de les encadrer entre deux nombres. Ils doivent notamment être en mesure de comparer deux fractions de même dénominateur et d'insérer un nombre décimal entre deux nombres entiers (CM1) puis deux nombres décimaux (CM2 et 6ème). Il est attendu qu'ils connaissent les égalités entre des fractions courantes.

La comparaison de nombres arabes apparaît alors comme une compétence fondamentale, enseignée et enrichie chaque année pour aider les élèves à approfondir leur compréhension des nombres et les préparer efficacement à des apprentissages plus avancés tels que le calcul et la résolution de problèmes mathématiques.

2.2. Aspects cognitifs : facteurs influençant la comparaison symbolique

La comparaison symbolique est une tâche complexe, faisant intervenir de nombreux mécanismes cognitifs. Ceux-ci peuvent affecter le traitement du nombre et interagir de différentes manières selon l'âge, le niveau de développement et les caractéristiques individuelles.

2.2.1. Dimensions numériques

Nous avons vu qu'au cours du développement, les représentations symboliques viennent se greffer sur la représentation analogique. Les effets de distance, de taille et de rapport, manifestés dans des tâches de comparaison analogique, sont ainsi reconduits sur des tâches de comparaison symbolique (Chazoule & Thevenot, 2018; Gilmore et al., 2007; Holloway & Ansari, 2009), bien que leur impact puisse varier en fonction du type de comparaison et du contexte. Ces effets sont en lien avec la perception des grandeurs numériques.

L'effet de distance correspond au fait que les humains sont plus précis et rapides pour comparer deux nombres éloignés (2 vs 7) que deux nombres proches (6 vs 7). Les enfants semblent affectés par cet effet dès l'âge de 5 ans : les temps de réaction et le nombre d'erreurs augmentent lorsque les quantités à comparer sont proches l'une de l'autre (Gilmore et al., 2007). L'effet de taille, quant à lui, traduit l'idée selon laquelle il est plus rapide de comparer des petits nombres (4 vs 8) que des grands nombres (42 vs 46), à écart équivalent. Quand la taille des deux nombres augmente, la difficulté à les comparer augmente de manière proportionnelle (Cohen Kadosh et al., 2008). Selon l'effet de rapport, plus le ratio de discrimination entre deux nombres est faible, plus il est difficile de les comparer. Au début du développement, le ratio de discrimination de l'enfant est de 3/4 à 3 ans, et de 5/6 à 6 ans. L'acuité numérique s'affine et permet à l'adulte de distinguer des quantités dont le rapport est d'environ 10/11 (Halberda & Feigenson, 2008).

Ces effets sont considérés comme des preuves de l'existence d'une ligne numérique mentale et de son rôle dans le traitement et la comparaison des quantités. Les nombres seraient positionnés dans l'espace selon une orientation gauche (plus petit) / droite (plus grand) (Dehaene et al., 1993). Cette orientation est mise en évidence par l'effet SNARC (Spatial Numerical Association of Response Code), qui se traduit en comparaison numérique par des performances meilleures lorsque le plus grand nombre est situé à droite de la paire. Par ailleurs, les représentations numériques mentales répondent à la loi de Weber-Fechner, c'est-à-dire qu'elles sont soumises à une compression logarithmique lorsque les quantités augmentent. Les effets observés ci-dessus montrent que les adultes obéissent à cette loi lorsqu'ils sont confrontés à une tâche de comparaison de quantités, et ce, quel que soit le format de présentation : symbolique ou non symbolique (Brannon, 2006).

Des tâches de comparaison symbolique peuvent également induire des effets d'interférence, en lien avec la congruence numérique.

Dans un contexte de comparaison de nombres à deux chiffres, on observe à partir de 7-8 ans, un effet de congruence selon lequel les sujets sont plus rapides à comparer deux nombres lorsque les chiffres des dizaines et des unités de chaque nombre coïncident, c'est-à-dire que leur comparaison conduit à la même décision (Girelli, et al., 2000; Rubinstein et al., 2002). Par exemple, ils sont plus rapides pour comparer 24 vs 99 que 48 vs 61.

Dans des tâches de comparaison impliquant des nombres décimaux ou des fractions, un biais des nombres naturels peut être observé : les enfants auraient tendance à transposer de manière incorrecte une propriété des nombres entiers (Vamvakoussi et al., 2012). Cela se traduit par de meilleurs résultats en situation congruente, qu'en situation incongrue dans ces deux types de comparaison. En comparaison de nombres décimaux, les enfants sont influencés par la propriété selon laquelle « plus le nombre de chiffres est grand, plus sa grandeur est grande » et font ainsi plus d'erreurs quand le plus grand nombre de la paire possède la plus petite partie décimale (Roell et al, 2017). Par exemple, ils peuvent juger que 0,32 est plus grand que 0,8 car « 32 » est plus grand que « 8 ». Dans la comparaison de fractions, les enfants, mais aussi les étudiants, les adultes et même les mathématiciens experts, sont plus rapides à donner des réponses correctes lorsque la fraction la plus grande contient les nombres naturels les plus grands que lorsqu'elle contient les nombres naturels les plus petits (Obersteiner et al., 2013; Vamvakoussi et al., 2012). Ils ont tendance à examiner séparément les composantes des nombres naturels qui constituent les fractions, et à commettre des erreurs de type « 4/5 est plus grand que 3/2, car 4 est plus grand que 3 ».

Les erreurs observées en situation d'incongruence numérique sont considérées comme des erreurs de raisonnement systématiques, selon la théorie du double processus du raisonnement humain. Il s'agit de la tendance à utiliser des stratégies heuristiques, automatiques, plutôt que des stratégies algorithmiques qui sont plus lentes et analytiques. Pour résister à cette tendance, un troisième système, celui de l'inhibition cognitive, doit intervenir (Houdé, 2018). Dans un contexte de comparaison de fractions et de nombres décimaux, il a notamment été démontré qu'un contrôle inhibiteur est essentiel pour supprimer les réponses intuitives automatiques, et garantir des réponses plus exactes (Borst et al., 2015; Roell et al., 2017, 2019).

2.2.2. Dimensions non numériques

Des recherches ont été menées afin d'étudier la manière dont les caractéristiques visuo-spatiales des symboles numériques peuvent influencer la comparaison symbolique et la prise de décision.

Des interactions entre le nombre et la grandeur physique ont été identifiées dans des tâches de comparaison symbolique. Les performances seraient moins bonnes en précision et en vitesse quand les paires sont incongruentes (Tzelgov et al., 1992). Autrement dit, cet effet de congruence de taille se traduit par des difficultés à comparer des nombres lorsque la grandeur numérique et la grandeur physique des nombres ne coïncident pas ; par exemple lorsque le chiffre 2 est affiché dans une police plus grande que le chiffre 8 (Schwarz & Ischebeck, 2003).

Des interférences s'observent également entre le nombre et la longueur physique lors de traitements arithmétiques plus complexes tels que la comparaison de nombres décimaux. Dans une étude menée auprès d'adolescents (12 ans) et de jeunes adultes (20 ans), les résultats ont montré l'impact de l'incongruence entre la grandeur des nombres et leur longueur physique sur la comparaison de nombres décimaux (Roell et al., 2019).

En contexte de comparaison symbolique, les individus peuvent être également influencés par la luminosité (ou la brillance) des nombres arabes. Des études ont mis en évidence que l'incongruence entre la luminosité et les nombres (par exemple, un chiffre sombre pour un nombre élevé et un chiffre lumineux pour un nombre faible), pouvait entraîner des temps de réaction plus lents et des erreurs plus nombreuses en comparaison de nombres arabes (Cohen Kadosh & Henik, 2006; Pinel et al., 2004).

Les études menées par Cohen Kadosh et Henik (2006) ainsi que Rubinsten et Henik (2005) ont également mis en évidence l'influence du contraste entre la figure (les nombres) et le fond (l'arrière-plan), sur la comparaison symbolique. Lorsque ce contraste est faible, les temps de réaction sont allongés et la précision réduite. En d'autres termes, lorsque les chiffres ne sont pas clairement distinguables de leur environnement en raison d'un faible contraste, les participants prennent plus de temps pour effectuer des comparaisons symboliques et font davantage d'erreurs.

L'ensemble de ces facteurs visuo-spatiaux joue un rôle important dans la manière dont les nombres sont perçus et comparés. Ces dimensions non numériques non pertinentes doivent être inhibées, afin de parvenir à comparer des nombres arabes de façon rapide et précise. Les effets d'interférence peuvent être plus importants chez les enfants car leurs mécanismes exécutifs ne sont pas aussi efficaces qu'à l'âge adulte (Dempster, 1992). Pendant l'âge scolaire notamment, l'attention sélective et l'inhibition d'informations visuelles s'améliorent progressivement, mais les enfants peuvent encore être sensibles aux distractions (Davidson et al., 2006).

3. Le Trouble des Apprentissages Mathématiques en lien avec la comparaison symbolique

Le Trouble des Apprentissages Mathématiques (TAM) est un trouble persistant et spécifique du développement numérique et des apprentissages mathématiques. Au cœur des difficultés qu'il représente, l'une des questions essentielles de notre étude réside dans la capacité à comparer des nombres arabes.

3.1. Critères diagnostiques du Trouble des Apprentissages Mathématiques

Selon la définition officielle du manuel diagnostique et statistique des troubles mentaux (American Psychiatric Association, 2013), le Trouble des Apprentissages Mathématiques (TAM) est un trouble du neurodéveloppement, persistant et spécifique, défini par des difficultés d'apprentissage et d'utilisation des capacités mathématiques. Parmi les symptômes, on identifie de manière générale, des difficultés à maîtriser le sens du nombre, les faits arithmétiques, le calcul, ainsi que des difficultés avec le raisonnement mathématique. Ces complications apparaissent chez des enfants en dépit d'une intelligence préservée, de l'absence de troubles sensoriels, neurologiques ou mentaux, d'un milieu socioculturel normalement stimulant et d'une scolarité adaptée. Se manifestant au cours de la scolarité, elles subsistent malgré les mesures et interventions mises en place (aménagements pédagogiques, rééducation orthophonique…). Ainsi, malgré des compensations possibles, les enfants qui présentent un TAM conservent un décalage par rapport aux enfants de leur âge. Par conséquent, le trouble impacte de façon significative la vie quotidienne et scolaire.

3.2. Hypothèses étiologiques

Différentes hypothèses cognitives tentent d'expliquer l'origine fonctionnelle du TAM. Selon les hypothèses cognitives numériques, le TAM résulterait d'un trouble du traitement du nombre. Dans l'intérêt de mieux comprendre les mécanismes sous-jacents, certains auteurs tentent de comparer les performances en comparaison symbolique avec celles en comparaison analogique. Deux postulats font ainsi débat dans la littérature.

Certains évoquent un déficit du sens du nombre (Butterworth, 2005; Von Aster & Shalev, 2007; Wilson & Dehaene, 2007). Le TAM serait directement lié à un déficit du traitement des représentations non symboliques du nombre et des représentations numériques mentales. Autrement dit, il s'agirait d'un trouble de la capacité à appréhender les quantités, entraînant des difficultés de comparaison, d'identification, d'estimation, de placement des nombres sur une ligne numérique, etc.

D'autres suggèrent un déficit d'accès au sens du nombre (Rubinsten & Henik, 2005). Cette hypothèse est notamment défendue par les résultats de plusieurs études qui montrent que certains enfants présentant un TAM, rencontrent des difficultés à comparer des codes symboliques arabes alors qu'ils parviennent à traiter les représentations non symboliques (De Smedt & Gilmore, 2011; Iuculano et al., 2008; Landerl & Kölle, 2009; Rousselle & Noël, 2007). Le TAM pourrait donc s'expliquer par des difficultés d'accès aux représentations mentales des quantités, à partir des nombres présentés symboliquement sous forme arabe.

3.3. Comparaison des performances des enfants avec TAM : tâches de comparaison symbolique vs. non symbolique

Les performances des enfants présentant un TAM en comparaison de nombres arabes peuvent varier en fonction des études, des méthodologies et des critères de définition de la dyscalculie.

Une tendance générale émerge dans la littérature. Des études impliquant des enfants âgés de 6 ans à 12 ans ont montré que la performance des enfants TAM est significativement plus pauvre que celle des enfants contrôles, en termes de latences et de taux d'erreurs, lorsqu'il s'agit de traiter la magnitude des nombres symboliques. Cette observation a été mesurée lors de comparaisons de paires de nombres à un chiffre (De Smedt & Gilmore, 2011; Desoete et al., 2012; Iuculano et al., 2008; Landerl et al., 2004; Landerl et al., 2009; Landerl & Kölle, 2009; Mussolin et al., 2010; Rousselle & Noël, 2007) et de paires de nombres à plusieurs chiffres (Desoete et al., 2012; Landerl et al., 2009; Landerl & Kölle, 2009).

D'autre part, une dissociation entre les performances en comparaison symbolique et les résultats en comparaison non symbolique est démontrée dans certaines études. Les observations diffèrent selon l'âge des enfants testés. Les études portant sur des enfants plus jeunes (6 à 9 ans), ont constaté des performances en comparaison symbolique plus faibles chez les enfants TAM, alors que leurs compétences en comparaison analogique sont similaires à celles des enfants contrôles (De Smedt & Gilmore, 2011; Iuculano et al., 2008; Landerl & Kölle, 2009; Rousselle & Noël, 2007). Les enfants TAM âgés de 10 ans et plus, sont quant à eux significativement moins performants en comparaison non symbolique que leurs pairs au développement typique (Mussolin et al., 2010; Price et al., 2007). Les difficultés de comparaison de collections se manifesteraient donc plus tardivement que les difficultés de comparaison symbolique chez les enfants TAM. Ces éléments sont plutôt en faveur d'un déficit d'accès au sens du nombre à partir des codes symboliques.

Alors que les enfants avec un TAM semblent présenter des performances significativement plus faibles que les enfants tout-venant du même âge, ils semblent également plus sensibles à l'effet de taille en comparaison symbolique de nombres à deux chiffres (Ashkenazi et al., 2009; Lander et al. 2009). D'autres recherches suggèrent que ces enfants peuvent être influencés par l'effet de distance entre deux nombres à comparer. Certaines mesurent une sensibilité accrue à cet effet (Ashkenazi et al, 2009; Mussolin et al, 2010; Price et al., 2007). Cela suggère une représentation mentale moins précise que chez les enfants tout-venant, entraînant une plus grande difficulté à traiter les relations de magnitude entre les nombres symboliques.

4. La comparaison de nombres arabes au sein d'une démarche diagnostique des TAM

4.1. Intérêt d'une épreuve de comparaison de nombres arabes

La comparaison numérique, qu'elle soit symbolique ou analogique, est une compétence mathématique fondamentale. La manière dont un individu aborde la comparaison symbolique par rapport à la comparaison analogique est un indicateur classique de sa compréhension des grandeurs numériques (Laski & Siegler, 2007). Selon Schwenk et al. (2017), les performances en comparaison de quantité symbolique seraient un marqueur central pour diagnostiquer un TAM de façon précoce. En considérant les données de la littérature concernant les performances plus faibles des enfants présentant un TAM que les sujets sains dans les tâches de comparaison de nombres arabes; administrer une telle épreuve apparaît indispensable. Celle-ci pourrait permettre de repérer l'existence d'un trouble cognitif numérique et de considérer la possibilité d'un diagnostic de dyscalculie.

À l'égard de l'hypothèse du déficit de l'accès au sens du nombre via les codes symboliques arabes, Lafay et al. (2014) recommandent l'utilisation de diverses épreuves permettant d'évaluer l'efficacité du SNA et du SNP, et les capacités d'accès à ces deux systèmes via les codes symboliques (arabe et oral). Une épreuve de comparaison de nombres arabes semble ainsi répondre à ces indications. En effet, une tâche de comparaison symbolique mesure l'intégrité du SNA, car elle permet d'analyser la capacité des individus à distinguer et estimer deux quantités numériques différentes, à partir de nombres arabes, sans utiliser de procédures de comptage ou de calcul précis. Une telle épreuve permet donc de mesurer la précision d'accès à la magnitude à partir du code arabe, ainsi que le transcodage entre les codes arabe et analogique. Cette tâche se révèle d'autant plus appropriée pour évaluer la rapidité avec laquelle les informations de magnitude sont accessibles à partir des codes symboliques, car elle permet de mesurer le temps d'exécution (De Smedt et al, 2009).

4.2. Liens entre les capacités à comparer les nombres arabes et les performances mathématiques

Plusieurs études ont permis de mettre en évidence l'existence d'un lien entre les compétences mathématiques et les capacités à comparer des nombres arabes. Des recherches menées auprès d'enfants âgés de 5 à 10 ans ont révélé que les performances en comparaison symbolique étaient corrélées aux scores obtenus à des tests de problèmes arithmétiques (Durand, et al., 2005; Hawes et al., 2019; Holloway & Ansari, 2009; Nosworthy et al., 2013). D'autres auteurs ont également montré que la performance en comparaison de nombres arabes chez les enfants de 6-7 ans était prédictive des compétences mathématiques un an plus tard (De Smedt et al., 2009, 2013; Mundy & Gilmore, 2009). Ces observations ont été constatées à travers des tests standardisés évaluant divers aspects mathématiques tels que la connaissance des nombres, la compréhension des opérations, l'arithmétique, la résolution de problèmes, la mesure et la géométrie. Certains chercheurs précisent que chez les jeunes enfants, le traitement de la magnitude sous forme symbolique s'avère être un meilleur indicateur des compétences mathématiques futures que le traitement de la magnitude analogique (Schneider et al., 2017). Il revêtirait même une importance pour prédire les compétences en mathématiques à l'école primaire, équivalente à celle des compétences phonologiques pour prédire les performances futures en lecture (Vanbinst et al., 2016).

La comparaison de nombres arabes est un marqueur de la compréhension du système de numération à valeur positionnelle (Moeller et al., 2011). Elle est nécessaire pour comprendre d'autres concepts tels que les différents types d'opérations pour les calculs. Jordan et al. (2009) ont montré que les compétences en comparaison de nombres étaient prédictives de la progression des compétences en calcul et en résolution de problèmes, de la première à la troisième année. La comparaison symbolique joue également un rôle dans les activités quotidiennes, car elle permet aux individus d'identifier les nombres, de les comprendre et de les utiliser de manière fonctionnelle (Lafay et al., 2013).

L'évaluation de la comparaison symbolique paraît donc intéressante pour diagnostiquer un TAM, car des difficultés en comparaison de nombres arabes peuvent indiquer des difficultés de compréhension des mathématiques en général. Les résultats obtenus à une telle tâche permettraient ainsi de prédire les capacités ultérieures en mathématiques.

5. Buts et hypothèses

Les études mentionnées précédemment ont contribué à la compréhension du développement typique de la comparaison des nombres arabes et au rôle crucial de cette compétence dans les performances mathématiques. Il est également établi que l'évaluation de cette compétence chez les enfants d'âge scolaire contribue au diagnostic précoce, nécessitant une épreuve standardisée.

Ce mémoire s'inscrit alors dans une série de mémoires d'orthophonie dont le projet consiste à élaborer une batterie d'évaluation diagnostique de la cognition mathématique chez les enfants d'âge scolaire. Nous participons à la validation de sept épreuves de cette batterie évaluant respectivement la comparaison de nombres arabes, les gnosies auditivo-verbales de nombres, le dénombrement, la chaîne numérique verbale, les faits arithmétiques, les problèmes verbaux et les problèmes analogiques.

Notre étude se concentre uniquement sur la phase de pré-test de l'épreuve de comparaison de nombres arabes, préalablement conçue à partir d'une analyse critique des épreuves de protocoles existants évaluant cet aspect (Cousin, 2021). Elle a pour objectif principal d'évaluer la pertinence et la qualité de mesure de cette épreuve, en passant par l'analyse et l'interprétation des résultats obtenus lors de premières passations. Les analyses quantitatives et qualitatives nous permettront de valider la phase de pré-test et d'identifier les ajustements nécessaires.

À travers ce travail initial de validation, nous souhaitons vérifier la bonne construction de l'épreuve, la clarté des instructions et la conformité des items aux performances attendues selon le niveau des enfants. Si cette hypothèse est vérifiée, nous observerons chez les sujets témoins sollicités pour la validation, des effets de niveau scolaire et de complexité des items.

Méthode

Dans cette partie, nous présenterons l'épreuve de comparaison de nombres arabes dont la création a fait l'objet du précédent mémoire (Cousin, 2021). Afin de poursuivre le travail entamé, nous avons effectué le pré-test de cette épreuve. Nous expliquerons alors notre protocole expérimental, qui comprend le recrutement d'une population d'enfants tout-venant, la passation de l'épreuve auprès de ces sujets, et les méthodes d'analyse des données recueillies.

1. Matériel

1.1. Création de l'épreuve de comparaison de nombres arabes

Le travail théorique et expérimental mené par Cousin (2021) au cours du précédent mémoire, a permis d'aboutir à la création d'une épreuve de comparaison de nombres arabes.

Innovante et exhaustive, il s'agit de la seule épreuve de comparaison de nombres arabes qui est à la fois chronométrée, qui comporte soixante items (dont des fractions et des nombres décimaux) et qui aspire à être informatisée et à intégrer des critères de début et d'arrêt. Le chronométrage de l'épreuve permet d'optimiser la fidélité inter-juge et de répondre ainsi à une bonne standardisation. La mesure des temps de réponse est d'autant plus importante car elle permet d'évaluer l'efficacité du processus de traitement des nombres. Les scores obtenus par le sujet à cette épreuve reflètent quant à eux sa précision numérique et la qualité du fonctionnement de son SNA. Enfin, la construction de l'épreuve a fait en sorte de faire varier certains critères, afin de mettre en évidence ce qui pose le plus de difficultés à l'enfant : la congruence des paires, l'ordre de la ligne numérique et le ratio.

L'épreuve a donc été créée dans l'objectif d'une évaluation plus fine et complète de l'accès au code arabe et au code analogique, ainsi que du transcodage entre ces deux codes.

1.2. Description de l'épreuve

Cette épreuve consiste à présenter au sujet, plusieurs paires de nombres arabes à comparer. Chaque paire constitue un item. Les items sont regroupés selon cinq catégories, construites en lien direct avec les étapes du développement des enfants, ainsi qu'avec les acquisitions scolaires de la maternelle (cycle 1) jusqu'au collège (cycle 3) :

- Une catégorie de petits nombres (de 1 à 4),
- Une catégorie de nombres moyens (de 4 à 14),
- Une catégorie de grands nombres (de 15 à 99),
- Une catégorie de très grands nombres (de 100 à 45 700),
- Une catégorie de fractions et de nombres décimaux.

Pour chacune de ces catégories, douze items sont affichés en noir sur un fond blanc, en respectant un ordre précis. L'ensemble de l'épreuve se compose donc au total de soixante items, et de deux items d'exemple. Les tableaux reprenant les items et leurs critères d'après Cousin (2021) sont disponibles en annexe 1.

1.3. Déroulement de l'épreuve

Dans le contexte actuel de pré-test, l'épreuve est présentée sous format « papier-crayon ». On présente au sujet cinq colonnes composées de douze rectangles chacune, représentant de manière respective les cinq catégories, et les douze items de paires de nombres à comparer. La consigne de l'épreuve est la suivante : « Pour chaque rectangle, barre le plus vite possible le plus grand des deux nombres ». Un exemple de deux paires à comparer est proposé à l'enfant avant le début de l'épreuve.

Le chronomètre est déclenché dès que l'examinateur a fini de lire la consigne, et est arrêté à la fin de la passation de l'épreuve. L'examinateur ne donne pas d'indication sur l'exactitude des réponses données par le sujet.

Les critères de début et d'arrêt envisagés par Cousin (2021) ne sont pas appliqués dans ce contexte de pré-test, afin de ne pas limiter notre exploration et notre capacité à identifier les ajustements nécessaires.

1.4. Cotation

La cotation est binaire pour chaque item : l'enfant obtient 1 point à l'item si sa réponse est correcte, et 0 si sa réponse est erronée ou s'il ne parvient pas à répondre.

Les temps de réponse sont calculés par l'examinateur. Nous pouvons avoir accès au temps de réponse par catégorie, et au temps de réponse total.

2. Population

2.1. Recrutement des sujets

Pour notre étude, nous avons fait le choix de constituer une population d'enfants tout-venant, scolarisés du CP au CM2. Le recrutement de cette population s'est réalisé dans les départements du Nord-Pas-de-Calais (62) et de la Haute-Garonne (31). Nous avons contacté une dizaine d'Inspecteurs de l'Education Nationale par courrier électronique, afin de leur exposer notre projet et solliciter leur autorisation d'intervenir dans les écoles de leurs circonscriptions. Après l'accord de trois d'entre eux, nous avons pu contacter une vingtaine d'établissements scolaires. Deux écoles de la Haute-Garonne ainsi que deux écoles du Nord-Pas-de-Calais se sont portées volontaires pour la réalisation de nos passations. La rencontre avec les équipes éducatives de ces écoles nous a permis de répondre aux éventuelles questions de nature organisationnelle, et de leur transmettre les autorisations parentales (cf. annexe 2). Le nombre d'enfants pour lesquels les parents ont accordé leur autorisation parentale était significativement élevé par rapport à la contrainte de temps disponible dans chaque école. Afin de garantir une sélection équitable, nous avons opté pour une approche aléatoire dans le choix des participants. Cette méthode a été employée pour obtenir une répartition aussi représentative que possible, tenant compte des critères tels que le sexe, l'âge, et le niveau scolaire.

2.2. Présentation de la population

Notre population se compose de 80 enfants tout-venant. Ces élèves, scolarisés du CP au CM2, ont été recrutés dans deux départements de France, au sein de quatre écoles élémentaires publiques, en milieu rural. Les écoles A et B et les écoles C et D sont localisées respectivement dans les départements du Nord-Pas-de Calais (62) et de la Haute-Garonne (31).

Le pré-test de notre épreuve de comparaison de nombres arabes a été réalisé sur un échantillon de 79 enfants. Par contrainte de temps, nous n'avons pas pu administrer cette épreuve à un enfant de CP de notre population initiale. Parmi notre échantillon, quatre enfants ont bénéficié d'un suivi orthophonique pour trouble d'articulation, trouble alimentaire pédiatrique et retard de langage. Trois enfants sont actuellement suivis en orthophonie pour un trouble spécifique du langage écrit. Nous avons décidé de ne pas exclure ces enfants de notre échantillon car les pathologies mentionnées ne sont pas liées aux compétences mathématiques évaluées.

À partir de l'échantillon retenu, nous avons constitué cinq groupes correspondant aux cinq niveaux scolaires sollicités pour l'étude. Les tableaux ci-dessous présentent les répartitions de ces cohortes en fonction de trois critères : le sexe, l'âge, et l'établissement scolaire.

Tableau 1. Répartition de l'échantillon en fonction du sexe, selon les cinq niveaux scolaires (effectif entre parenthèses).

Niveau scolaire (effectif)	**CP (13)**	**CE1 (11)**	**CE2 (23)**	**CM1 (17)**	**CM2 (15)**	**Total (79)**
Ratio filles / garçons	9/4	4/7	11/12	8/9	9/6	41/38

Tableau 2. Répartition de l'échantillon en fonction de l'âge (âges minimum, maximum, moyen), selon les cinq niveaux scolaires.

	Age minimum	**Age maximum**	**Age moyen**
CP	6 ans 1 mois	6 ans 9 mois	6 ans 4 mois
CE1	7 ans	7 ans 8 mois	7 ans 4 mois
CE2	7 ans 10 mois	8 ans 10 mois	8 ans 4 mois
CM1	8 ans 4 mois	10 ans 4 mois	9 ans 5 mois
CM2	9 ans 7 mois	10 ans 10 mois	10 ans 4 mois

Tableau 3. Répartition de l'échantillon en fonction des établissements scolaires, selon les cinq niveaux scolaires.

	Ecole A	**Ecole B**	**Ecole C**	**Ecole D**	**Effectif total**
CP	7	3	4	0	13
CE1	1	2	4	4	11
CE2	14	0	3	6	23
CM1	6	3	4	4	17
CM2	4	3	4	4	15
Effectif total	31	11	19	18	79

3. Méthode

3.1. Organisation et déroulement des passations

Nous sommes deux étudiantes à avoir réalisé un total de 80 passations dans deux écoles de la Haute-Garonne (31) et deux écoles du Nord-Pas-de-Calais (62), sur une période d'un mois.

Nous avons veillé à ce que l'épreuve se déroule dans une pièce calme de l'établissement scolaire, afin que les conditions de concentration et d'attention soient les plus optimales possibles. Chaque passation s'est donc réalisée au cours de séances uniques, en situation duelle, dans une salle spécialement mise à disposition par les directrices d'écoles.

Une semaine avant notre intervention dans les écoles, les dates et les objectifs de notre expérimentation avaient été communiqués aux élèves par leurs professeurs. Avant chaque évaluation et de manière systématique, nous avons toutefois tenu à rappeler l'objectif de l'étude, l'anonymat du test et le déroulement de la passation. Cet entretien préalable a également permis de répondre aux éventuelles questions, de vérifier l'autorisation écrite des parents mais aussi de s'assurer du consentement de l'enfant et de confirmer les données personnelles nécessaires à notre analyse statistique.

Nous avons ensuite administré les sept épreuves de la batterie d'évaluation selon un ordre randomisé. L'épreuve de comparaison de nombres arabes a été présentée sous format « papier-crayon ». Deux items d'exemple ont été proposés au sujet pour le familiariser avec la tâche à accomplir et le mettre en confiance. Chaque participant s'est vu administrer les soixante items de l'épreuve et a pu noter ses réponses directement sur le protocole. Les enfants étaient libres d'interrompre l'épreuve à n'importe quel moment. La durée maximale allouée à l'épreuve de comparaison de nombres était de 9 minutes. Au total, chaque passation a duré entre 20 et 30 minutes par enfant.

3.2. Méthodes d'analyse des réponses

La cotation des épreuves s'est effectuée lors d'un temps spécifique. Par souci d'anonymisation, les données recueillies (scores et temps) ont été associées à un numéro propre à chaque enfant, puis extraites sous forme de tableur Excel. Ce dernier fait donc apparaître, pour chaque enfant, le score par item, le score et le temps par catégorie, le score total et le temps total.

Dans l'objectif de vérifier la bonne construction et la faisabilité de notre épreuve, nous avons étudié les scores moyens et les écarts-types de notre population, ainsi que les temps moyens de réalisation. Pour ces derniers, certaines données n'étaient pas interprétables en raison de plusieurs limites identifiées : 3 élèves ont abandonné l'épreuve à partir de la quatrième catégorie d'items, 32 enfants n'ont pas réalisé la cinquième catégorie et 7 élèves ne l'ont complétée que partiellement. Par conséquent, ces 42 enfants ont obtenu des temps de réalisation peu élevés, voire nuls, pour ces deux catégories d'items, ne reflétant pas leurs compétences réelles. Afin d'interpréter les temps moyens tous items confondus, seules les données des enfants ayant terminé l'épreuve ont été prises en compte (37 au total). L'analyse des temps moyens en fonction des catégories d'items a été réalisée à partir des données de nos 79 sujets pour les trois premières catégories, des données de 76 d'entre eux pour la catégorie 4, puis des données des 37 enfants ayant terminé l'épreuve pour la catégorie 5.

Afin de sélectionner les items qui assureront au mieux la validité de l'épreuve, une analyse d'items classique a été réalisée, tous niveaux scolaires confondus, puis en fonction des classes. Dans l'approche psychométrique classique, les items sont principalement caractérisés par deux mesures :

leur niveau de difficulté et leur indice de discrimination (Demeuse & Henry, 2004). L'indice de difficulté p d'un item (compris entre 0 et 1), est le rapport entre le nombre de sujets qui répondent correctement à l'item et le nombre de personnes soumises au test. Plus l'indice p est élevé, plus l'item est facile. Un test doit comporter des items de niveaux de difficulté variés, mais il est recommandé que la majorité des items possède un indice de difficulté variant entre 0,3 et 0,7 (Laveault & Grégoire, 2023). Par ailleurs, les items sont généralement classés selon leur niveau de difficulté (Chartier & Vrignaud, 2018). L'indice de discrimination D d'un item (également compris entre 0 et 1) évalue la capacité d'un item à discriminer les sujets forts des sujets faibles, en comparant leur réussite à l'item. Le groupe fort regroupe les 27% de participants ayant obtenu les scores les plus élevés à l'épreuve concernée, tandis que le groupe faible rassemble les 27% avec les résultats les plus faibles. L'indice D d'un item est le résultat de la différence entre la moyenne du groupe fort et la moyenne du groupe faible à cet item. Plus la valeur de l'indice est proche de 1, plus le pouvoir de différenciation de l'item est élevé (Laveault & Grégoire, 2023). Pour l'interprétation de nos résultats, nous nous appuierons sur la classification en cinq catégories d'Ebel (1965) : item sans utilité réelle pour l'examen (D < 0,10), item limite, à améliorer (D = 0,10 à 0,19), item qui discrimine peu (D = 0,20 à 0,29), item qui discrimine bien (D = 0,30 à 0,39), item qui discrimine très bien ($D \geq$ 0,40).

Résultats

Dans cette partie, il s'agira de présenter l'analyse qualitative et statistique des résultats issus de notre recueil de données.

1. Analyse qualitative des passations

Tous les élèves participant à notre étude ont accepté de réaliser le test. Aucun d'entre eux n'a sollicité d'explications supplémentaires après la présentation de la consigne, et aucun n'a échoué aux items exemples. Bien que la plupart des enfants aient compris la consigne, quatre d'entre eux n'ont pas respecté une instruction : au lieu de barrer les nombres les plus grands parmi chaque paire, ils les ont entourés. Nous relevons une variabilité dans la modalité de réponses pour les autres enfants : certains ont fait le choix de barrer d'un seul trait, tandis que d'autres ont préféré réaliser une croix.

La volonté de réussir l'épreuve a été démontrée par la majorité des enfants. Certains ont pris leur temps pour répondre avec précision, d'autres ont privilégié la rapidité d'exécution. Cependant, 44 élèves ont rencontré des difficultés lors de la réalisation de la dernière catégorie d'items évaluant la comparaison de fractions et de nombres décimaux. Parmi eux, 32 ont fait le choix d'abandonner l'épreuve à partir de cette catégorie, car elle impliquait des nombres qu'ils n'avaient pas encore étudiés (69% des CP, 64% des CE1, 57% des CE2 et 18% des CM1). Les douze restants ont mentionné leur volonté d'essayer ou de deviner les bonnes réponses. Seuls cinq d'entre eux ont répondu à l'ensemble des items de la catégorie. Parmi eux, deux enfants ont effectué des comparaisons entre les parties d'un même nombre (les numérateurs et les dénominateurs d'une même fraction, ou les parties entière et décimale d'un même nombre décimal). De plus, nous comptabilisons trois abandons à partir de la quatrième catégorie par des élèves de CP, et l'apport de deux aides dans cette même catégorie pour un autre enfant.

2. Analyse quantitative des résultats des passations

Les méthodes d'analyse quantitative décrites précédemment nous ont permis d'obtenir des données statistiques que nous rapportons dans cette section.

2.1. Scores moyens et écarts-types selon les différents niveaux scolaires

2.1.1. Scores moyens et écarts-types tous items confondus

Le tableau ci-dessous présente les moyennes et les écarts-types obtenus par les cinq niveaux scolaires, à l'épreuve de comparaison de nombres arabes. La moyenne représente le centre des données, alors que l'écart-type nous communique des informations sur l'étalement des données autour de la moyenne. Plus l'écart-type est faible, plus la population est homogène.

Tableau 4. Scores moyens (sur 60), minimum et maximum, et écarts types obtenus à l'épreuve de comparaison de nombres arabes, tous items confondus, en fonction des cinq niveaux scolaires.

	CP	**CE1**	**CE2**	**CM1**	**CM2**
Moyenne (sur 60)	40	49	50	54	59
Score min - Score max	[21 – 48]	[43 – 55]	[42 – 58]	[48-59]	[55-60]
Ecart-type	7,81	4,43	5,13	3,79	1,39

Les résultats rapportés dans le Tableau 4 montrent une augmentation des moyennes en fonction des niveaux scolaires. On relève cependant une stabilité plus prononcée entre le CE1 et le CE2.

Par ailleurs, on note une diminution globale de la valeur des écarts-types en fonction des niveaux scolaires, illustrée par la Figure 1. Cette dernière montre en effet une différence de distribution des scores entre les classes : les performances des CM2 sont plus homogènes.

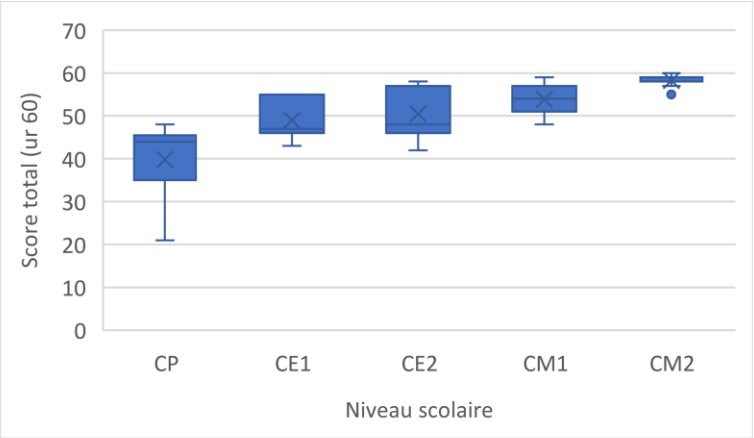

Figure 1. Boîtes à moustaches des scores tous items confondus (sur 60) obtenus à l'épreuve de comparaison de nombres arabes, en fonction des cinq niveaux scolaires.

2.1.2. Scores moyens en fonction des catégories d'items

Le tableau ci-dessous permet d'analyser les scores moyens par catégorie d'items (sur 12), en fonction des cinq niveaux scolaires.

Tableau 5. Scores moyens par catégorie d'items (sur 12) obtenus à l'épreuve de comparaison, en fonction des cinq niveaux scolaires.

	CP	CE1	CE2	CM1	CM2
Catégorie 1	11,85	12	11,96	12	12
Catégorie 2	11,62	12	12	11,94	12
Catégorie 3	10	12	11,87	11,59	12
Catégorie 4	5,85	10,36	10,96	11,53	11,73
Catégorie 5	0,46	2,64	3,7	6,82	10,53

Les scores moyens des trois premières catégories, sont élevés et relativement stables pour chaque niveau scolaire. Des effets plafonds sont mêmes observés dès la première catégorie pour les CE1, CM1 et CM2 (tous les élèves obtiennent de très bons scores). Concernant les catégories d'items les plus avancées (catégories 4 et 5), les scores moyens diminuent pour tous les niveaux. Ils tendent à être plus bas chez les élèves de CP, CE1 et CE2 par rapport aux niveaux supérieurs (CM1 et CM2). Les scores moyens de CM2 sont quant à eux les plus élevés pour ces mêmes catégories.

2.2. Temps moyens selon les différents niveaux scolaires et tous niveaux confondus

2.2.1. Temps moyens tous items confondus

La figure ci-dessous présente la distribution des temps de réalisation tous items confondus (en secondes), des 37 enfants ayant complété l'intégralité de l'épreuve, en fonction des niveaux scolaires puis tous niveaux confondus.

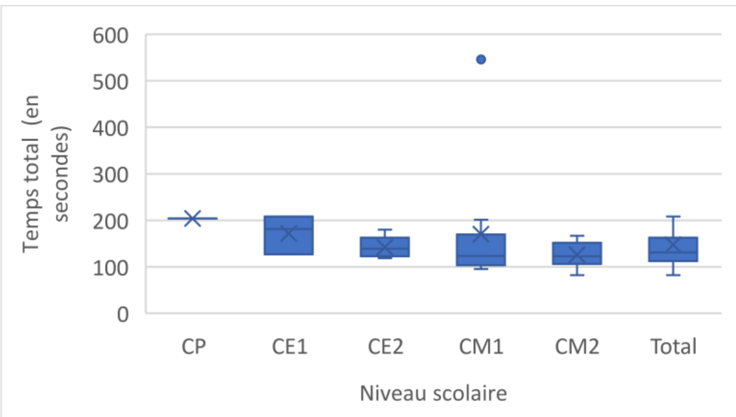

Figure 2. Boîtes à moustaches des temps de réalisation tous items confondus (en secondes) obtenus à l'épreuve de comparaison de nombres arabes, en fonction des cinq niveaux scolaires, et tous niveaux confondus.

Dans un premier temps, nous pouvons observer que les élèves de CP affichent le temps de réalisation moyen le plus élevé (204 secondes), tandis que les élèves de CM2 possèdent le temps de réalisation moyen le plus bas (126 secondes). Nous notons cependant une variation inattendue chez les élèves de CM1, avec un temps moyen plus élevé par rapport aux élèves du niveau inférieur. Cette variation est notamment marquée par une valeur extrême chez les élèves de CM1, correspondant à un temps maximal de 546 secondes.

Par ailleurs, la Figure 2 nous donne une indication importante sur la durée de l'épreuve. En moyenne, 147 secondes soit 2 minutes et 27 secondes ont été nécessaires pour que les élèves terminent le test.

2.2.2. Temps moyens en fonction des catégories d'items

Nous avons étudié les temps moyens de réalisation par catégorie d'items, en fonction des différents niveaux scolaires puis tous niveaux scolaires confondus. Le tableau ci-dessous présente les résultats obtenus.

Tableau 6. Temps moyens par catégorie d'items (en secondes) obtenus à l'épreuve de comparaison de nombres arabes, en fonction des cinq niveaux scolaires et tous niveaux confondus (effectif des sujets ayant terminé la catégorie entre parenthèses).

Niveau scolaire (effectif)	CP (79)	CE1 (79)	CE2 (79)	CM1 (76)	CM2 (37)	Total
Catégorie 1	28	14	15	13	11	16
Catégorie 2	32	21	19	15	13	19
Catégorie 3	58	25	25	21	17	28
Catégorie 4	67	52	38	28	27	39
Catégorie 5	120	62	48	91	59	63

Une première analyse des résultats du Tableau 6 permet de remarquer que pour chaque niveau scolaire, les catégories 4 et 5 sont les plus longues à réaliser tandis que les catégories 1, 2 et 3 sont les plus courtes. Les élèves de CP ont tendance à avoir des temps de réalisation plus longs dans les catégories 4 et 5 par rapport aux autres niveaux scolaires, tandis que les élèves de CM2 ont des temps de réalisation plus courts dans toutes les catégories. Nous retrouvons une particularité pour le groupe d'enfants scolarisés en CM1, dont le temps moyen pour la catégorie 5 est plus élevé que pour les CE2. Les élèves de CE2 apparaissent également plus rapides que les CM2 pour réaliser cette dernière catégorie d'items.

2.3. Indices de difficulté (*p*) et de discrimination (*D*) par item

L'indice de difficulté *p* et l'indice de discrimination *D* ont été calculés pour chacun des items de l'épreuve de comparaison de nombres arabes. Les calculs ont été réalisés tous niveaux scolaires confondus, puis selon les cinq niveaux scolaires. Les résultats sont exprimés dans un tableau en annexe 3.

2.3.1. Indices *p* et *D* tous niveaux scolaires confondus

Pour une lecture plus rapide et plus lisible des résultats, nous présentons ci-dessous un graphique de type radar (cf. Figure 3), associant les indices de discrimination et de difficulté par item, tous niveaux confondus.

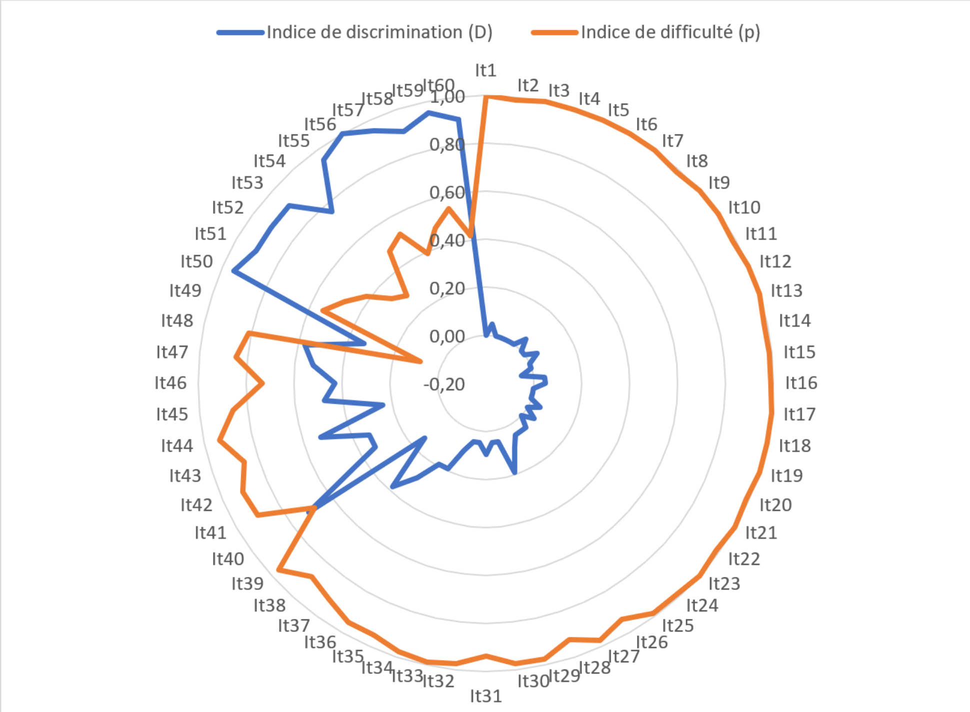

Figure 3. Indices de difficulté (*p*) et indices de discrimination (*D*) par item, obtenus à l'épreuve de comparaison de nombres arabes, tous niveaux scolaires confondus.

L'analyse conjointe des indices de difficulté et de discrimination des items (cf. Figure 3) permet de constater plusieurs points :

- 11 items rencontrent une difficulté moyenne ($30 \leq p \leq 70$) et une bonne discrimination ($D > 0,3$) : items 40 (catégorie 4), 50 à 53 et 55 à 60 (catégorie 5).
- 49 items possèdent un indice de difficulté situé dans des zones de discussion :
    - 1 item rencontre une très grande difficulté ($p < 0,10$) mais une bonne discrimination ($D > 0,3$) : item 49 (catégorie 5)
    - 1 item possède un indice *p* dans la zone [0,10 ; 0,30[et un très bon indice de discrimination ($D > 0,40$) : item 54 (catégorie 5)
    - 7 items ont un indice *p* dans la zone]0,70 ; 0,90] et un très bon indice de discrimination ($D > 0,40$) : items 38, 41, 43, 45, 46, 47, 48 (catégorie 4)
    - 40 items rencontrent une très grande facilité ($p > 0,90$)
        - 39 items n'ont pas un bon pouvoir discriminant ($D < 0,3$) : items 1 à 36 (catégories 1, 2, 3), items 37, 39, 44 (catégorie 4)
        - 1 item discrimine bien ($D > 0,3$) : 42 (catégorie 4)

Finalement, la Figure 3 illustre de manière générale qu'un item dont l'indice de difficulté est élevé ne possède pas nécessairement de qualité discriminante. Nous observons en effet que plus un item est très bien réussi par la majorité des élèves, moins il est efficace pour distinguer les élèves performants des élèves moins performants, et inversement. Il est également visible que l'ordre de présentation des items n'est pas complètement corrélé avec le niveau de difficulté des items. Ce classement nous donne des indications sur la construction de notre épreuve, que nous détaillerons dans la discussion.

2.3.2. Indices *p* et *D* en fonction des niveaux scolaires

À partir des indices calculés en fonction des niveaux scolaires (cf. annexe 3), nous avons pu dresser un classement des items en fonction des indices de difficulté (*p*), pour chaque niveau scolaire. Sont notés en gras, les items pour lesquels les indices de discrimination (*D*) sont inférieurs à 0,20.

Tableau 7. Classement des items de l'épreuve de comparaison de nombres arabes, en fonction des indices de difficulté et de discrimination (en gras) pour chaque niveau scolaire.

	CP	**CE1**	**CE2**	**CM1**	**CM2**
$p < 0,10$	**49**, 50, 51, **52**, **53**, 54, 55, **56**, **57**, **58**, 59 60	**49**, 53, **54**			
$0,10 \leq p < 0,30$	40	51, 55, 57, 60	**49**, 53, 54	49	49
$0,30 \leq p \leq 0,70$	28, 35-38, 41-48	40, 50, 52, 56, 58, 59	40, 50-52, 55-60	51-54, 57, 58, 60	
$0,70 > p \geq 0,90$	26, 27, 31, 34, 39	**38**, 43, 46, 48	45, 46, 48	28, 40, **46**, 48, 50, 55, 56, 59	46, **53, 54, 57, 60**
$p > 0,90$	**1, 2, 3, 4, 5, 6, 7**, 8, **9, 10, 11, 12, 13, 14**, 15, 16, **17, 18, 19**, 20, **21**, 22, **23**, 24, 25, 29, 30, **32**, **33**	**1-36**, 37, **39**, **41**, **42**, **44**, 45, 47	**1-39**, **41**, **42**, 43, **44**, **47**	**1-25**, 26, **27**, **29-39**, **41**, **42, 43, 44**, 45, **47**	**1-45,** 47, **48**, **50**, **51**, 52, **55**, **56**, 58, **59**

Note. *p* : indice de difficulté.

Comme l'illustre le Tableau 7, l'analyse des indices de difficulté par item pour chaque classe permet de confirmer que les élèves de tous les niveaux réussissent les items 1 à 39 (indices $p > 0,30$). Nous remarquons cependant que les élèves de CP montrent des niveaux de compétence variables à partir de l'item 26. La majorité d'entre eux échoue particulièrement aux items 49 à 60, pour lesquels les indices de difficulté inférieurs à 0,10 témoignent d'un effet plancher (tous les élèves de CP obtiennent des scores très faibles). Parmi ces items échoués, la moitié ne possède pas une bonne qualité discriminante (indice $D < 0,19$). Les items de la catégorie 5 sont également moins bien réussis par les élèves de CE1 et CE2. Les indices de difficulté sont inférieurs à 0,30 pour 7 items sur 12.

Seuls les items 49 et 54 ne sont pas discriminants. L'usage de cette catégorie d'items pour ces trois niveaux scolaires sera discuté.

Pour les élèves de CM1 et de CM2, les items 1 à 48 obtiennent des pourcentages de réussite entre 70% et 100%, témoignant la facilité pour ces enfants à répondre aux items des quatre premières catégories (effet plafond). Leurs performances se distinguent davantage à partir de la dernière catégorie d'items. Alors que les élèves de CM2 montrent une maîtrise apparente de ces items plus complexes, les élèves de CM1 obtiennent des résultats hétérogènes.

Enfin, nous pouvons identifier l'item 49 comme l'item le moins bien réussi par l'ensemble des élèves (indices $p < 0{,}30$). Son pouvoir discriminatif est nul pour les classes de CP, CE1 et CE2.

Discussion

Après avoir rappelé les objectifs et les hypothèses de notre étude, nous procèderons dans cette partie à l'interprétation des résultats obtenus, tout en identifiant les limites de notre travail et en proposant des pistes de développement pour de futures études.

1. Rappel des objectifs et hypothèses

Dans le cadre d'un projet de création d'une batterie d'évaluation diagnostique de la cognition mathématique, notre étude a pour objectif de procéder à l'étape initiale de validation de l'épreuve de comparaison de nombres arabes. Nous avons ainsi réalisé le pré-test de l'épreuve créée par Cousin (2021), auprès d'enfants tout-venant, au sein d'écoles primaires. Les passations effectuées ont permis d'obtenir les premières données qualitatives et quantitatives, nous permettant de vérifier la bonne construction de l'épreuve et la pertinence de nos items, et d'établir dans le cas contraire des modifications.

L'épreuve de comparaison de nombres arabes a été élaborée en tenant compte d'une progression de complexité. Les items sont regroupés selon cinq catégories, en accord avec les étapes du développement cognitif des enfants tout-venant et leurs acquis scolaires (Cousin, 2021). Nous nous attendons donc à observer des effets de niveau scolaire et de complexité des catégories d'items, qui confirmeraient la pertinence et la bonne construction de notre épreuve.

2. Interprétation des résultats

Nous allons à présent interpréter les résultats rapportés dans la partie précédente, afin de vérifier nos hypothèses.

2.1. Analyse des résultats principaux

2.1.1. Scores moyens

L'analyse des premiers résultats nous permet de constater une augmentation des moyennes en fonction des niveaux scolaires, chez les sujets témoins sollicités. Cette observation témoigne d'une tendance générale à l'amélioration des performances en comparaison de nombres arabes, à mesure que les élèves progressent dans leur scolarité. Entre le CE1 et le CE2, les compétences sont plus homogènes ou stationnaires. Ce résultat peut être un effet réel, ou bien la conséquence de la petite cohorte CE1, ou encore la conséquence de la construction de notre épreuve. Une vérification devra être réalisée avec une cohorte mieux dispersée et plus importante afin d'objectiver si c'est notre petite cohorte qui a biaisé nos résultats ou non. Une diminution de la valeur des écarts-types en fonction des niveaux scolaires est par ailleurs relevée, signifiant que plus le niveau scolaire des élèves est élevé, plus les performances des élèves tendent à se rapprocher de la moyenne. La plus grande homogénéité des performances est observée chez les élèves de CM2, où l'écart-type est le plus bas. Ce résultat suggère que ces élèves ont des performances moins variables dans la comparaison de nombres arabes par rapport aux élèves des autres niveaux scolaires. Cela pourrait être en lien avec la consolidation des apprentissages et l'approfondissement des enseignements mathématiques (Ministère de l'Education Nationale et de la Jeunesse, 2020). Cette progression générale permet finalement de confirmer les effets de niveau scolaire.

Afin de vérifier les effets de complexité des catégories d'items, nous avons examiné les scores moyens répartis selon les cinq catégories d'items et les cinq niveaux scolaires. Ces catégories suivent une progression ascendante selon la taille et la nature des nombres à comparer, allant des nombres arabes de 1 à 4, de 4 à 14, de 15 à 99, de 100 à 45 700, jusqu'aux nombres décimaux et aux fractions. On remarque que les élèves de CP semblent aussi bien réussir que les élèves de CM2 à comparer les nombres arabes des trois premières catégories, évaluant la comparaison des nombres allant jusqu'à 99. Cela est en accord avec les enseignements mathématiques, réalisés à partir de nombres inférieurs ou égaux à 100 au CP (Ministère de l'Education Nationale et de la Jeunesse, 2020). Une diminution progressive des performances est cependant observée à partir de la catégorie 4 (nombres compris entre 100 et 45 700). Les élèves de CP, CE1 et CE2 montrent les scores moyens les plus bas pour la catégorie 5 composée de fractions et de nombres décimaux, alors que les élèves de CM1 et CM2 obtiennent les scores les plus élevés. Cela rejoint l'analyse qualitative selon laquelle 69% des CP, 64% des CE1, 57% des CE2 et 18% des CM1 ont abandonné l'épreuve à partir de cette catégorie, car elle impliquait des nombres qu'ils n'avaient pas encore étudiés. Ces données qualitatives et quantitatives sont conformes aux attentes, car selon le bulletin officiel de l'Education nationale, les fractions et les nombres décimaux ne sont enseignés qu'à partir du CM1 (Ministère de l'Education Nationale et de la Jeunesse, 2020).

Ces résultats reflètent la différence des attendus académiques selon les niveaux scolaires, définis par les programmes officiels de l'Education nationale (Ministère de l'Education Nationale et de la Jeunesse, 2020). Leur analyse permet ainsi de vérifier que les performances évoluent avec le niveau scolaire et la complexité des catégories d'items. Les scores moyens ne sont pas homogènes et diffèrent les uns des autres comme nous l'avions espéré.

2.1.2. Temps moyens

En ce qui concerne les temps moyens de réalisation, il aurait été attendu de constater une diminution progressive de manière homogène avec les scores moyens relevés, indiquant une progression dans la maîtrise de la comparaison de nombres arabes en fonction de l'avancée dans la scolarité. Cependant, les élèves de CM1 sont plus lents que les élèves de CE2. Comme l'illustre la Figure 2, ce résultat est en partie attribuable à un cas particulier au sein des élèves de CM1, pour lequel une durée de 9 minutes environ a été nécessaire pour terminer l'épreuve. Cela semble avoir influencé la moyenne globale, entraînant un temps moyen plus élevé pour cette catégorie d'élèves. La comparaison des temps de réalisation moyens entre les CP et les CM2 permet toutefois de relever que ces derniers sont à la fois plus performants et plus rapides en comparaison de nombres arabes, que les élèves en début de cycle élémentaire.

L'analyse plus approfondie des temps moyens par catégorie montre également une corrélation entre la nature de la catégorie et le temps de réalisation : plus la catégorie est complexe, plus les temps de réalisation augmentent. Cependant, des variations dans les temps de réalisation existent entre les niveaux scolaires en fonction des catégories d'items. Alors que les élèves de CM2 sont plus rapides dans toutes les catégories, les élèves de CP montrent des temps de réalisation plus longs dans les catégories 4 et 5. L'enseignement des nombres supérieurs à 100, des nombres décimaux et des fractions ne faisant pas partie du programme scolaire de CP, il est normal que ces dernières catégories d'items posent plus de difficultés à ces élèves, et nécessitent donc plus de temps pour être complétées (Ministère de l'Education Nationale et de la Jeunesse, 2020). À mesure que les élèves grandissent, les temps moyens par catégorie diminuent. Cela indique une plus grande efficacité dans le traitement des paires de nombres en fonction de la consolidation des acquis mathématiques. Nous retrouvons cependant quelques particularités. Les élèves de CE2 sont plus rapides que les élèves de CM1 et CM2 pour réaliser les items de la catégorie 5. La possibilité de réponses aléatoires introduit une source d'incertitude dans l'interprétation de ce temps de réalisation, car certains élèves pourraient avoir pris moins de temps en répondant au hasard. Par conséquent, ces résultats peuvent ne pas refléter de manière précise la réalité des performances de ces élèves dans la comparaison de nombres arabes.

Ces résultats principaux laissent supposer que l'épreuve de comparaison de nombres arabes est en adéquation avec les étapes d'acquisition des compétences mathématiques.

2.2. Analyse par item

Dans l'objectif de vérifier que les items sont appropriés pour l'évaluation, nous avons procédé au calcul de leurs indices de difficulté et de discrimination, tous niveaux confondus. Ces données ont été complétées par une analyse de ces indices en fonction des niveaux scolaires. Nous avons décidé d'interpréter les résultats en fonction des cinq catégories d'items.

2.2.1. Items des catégories 1, 2 et 3 (nombres compris entre 1 et 99)

Une première analyse montre que les items des catégories 1, 2 et 3 sont très faciles. Les items 1 à 36 obtiennent en effet un taux de réussite supérieur à 90%, tous niveaux confondus. Alors que les élèves de CP peuvent montrer des niveaux de compétences variables à partir de de la troisième catégorie, ceux du CE1 au CM2 réussissent à comparer sans difficulté les nombres arabes des trois premières catégories. Des effets plafonds sont également relevés dès le CP, où la grande majorité des élèves réussit parfaitement 16 items des deux premières catégories. La facilité des deux premières catégories pour tous les niveaux scolaires est attendue, car celles-ci évaluent les comparaisons de nombres arabes allant de 1 à 4, puis de 4 à 14, qui selon les programmes scolaires, sont abordées dès la maternelle (Ministère de l'Education Nationale et de la Jeunesse, 2020). La catégorie 3, quant à elle, évalue la capacité des élèves à comparer des nombres arabes compris entre 15 et 99. Etant donné que l'apprentissage de ces nombres est une partie intégrante du programme de CP (Ministère de l'Education Nationale et de la Jeunesse, 2020), il semble normal d'observer des différences de performances chez les élèves de ce niveau, à partir de cette catégorie. Cette variabilité est observée plus précisément à partir de l'item 26 (41 vs 32), pour lequel plusieurs facteurs entrent en jeu, notamment l'absence de congruence numérique, la présentation des nombres dans l'ordre inverse de la ligne numérique mentale et la petite taille du ratio.

La conservation des items de ces trois premières catégories semble finalement devoir se décider en fonction des indices de discrimination et en fonction des niveaux scolaires.

Les items 1 à 24 des deux premières catégories (paires de nombres compris entre 1 à 4 et entre 4 à 14), conjuguent un indice de difficulté (*p*) supérieur à 0,90 et un indice de discrimination (*D*) inférieur à 0,20 pour les élèves des classes de CE1 au CM2. Ces items trop faciles et peu discriminants ne semblent donc pas pertinents pour ces niveaux scolaires (Ebel, 1965; Laveault & Grégoire, 2023). En ce qui concerne les élèves de CP, 8 items sur 24 ont un pouvoir discriminant satisfaisant. Conserver les items simples des catégories 1 et 2 pour les élèves de CP pourrait permettre une meilleure distinction des performances des sujets faibles d'une part, et pourrait aider à instaurer un sentiment de confiance en début d'épreuve d'autre part.

Concernant la troisième catégorie (paires de nombres compris entre 15 et 99), la majorité des items démontre un bon pouvoir discriminant pour les élèves de CP, et un pouvoir discriminant nul pour les autres niveaux scolaires. Seuls les items 26 et 28 parviennent à distinguer les performances des élèves de CM1. Il serait prudent de vérifier les résultats auprès d'une cohorte plus grande afin de décider de leur pertinence pour les élèves de CE1, CE2, CM1 et CM2.

Finalement, malgré leurs indices de difficulté presque identiques tous niveaux scolaires confondus, les items de ces trois catégories ne suivent pas totalement un ordre de difficulté croissant, ce qui justifie la nécessité d'une révision de leur classement.

2.2.2. Items de la catégorie 4 (nombres compris entre 100 et 45 700)

Au sein de la catégorie 4 qui évalue la capacité des sujets à comparer des nombres compris entre 100 et 45 700, des différences sont observées, tous niveaux confondus.

Tout d'abord, l'item 40 est le seul à obtenir un indice de difficulté moyen tous niveaux confondus, comparativement aux autres items qui obtiennent un indice de difficulté élevé voire très élevé. Alors que la majorité des CP échoue à cet item, la plupart des CM2 le réussissent. L'item 40 (45 658 vs 45 700) est le seul item composé de deux nombres arabes à 5 chiffres, supérieurs à 10 000. Il est donc normal d'observer une augmentation progressive du taux de réussite en fonction des niveaux scolaires. Par ailleurs, cet item possède un bon pouvoir discriminant pour les élèves de CP, CE1, CE2 et CM1.

Bien que les items 38, 41, 43, 45, 46, 47 soient généralement bien réussis par la majorité des élèves tous niveaux confondus, leur indice de difficulté est légèrement plus faible que celui des items 37, 39, 42 et 44 de la même catégorie. Cette différence peut s'expliquer par un effet de congruence. Les chiffres des unités, dizaines ou centaines des paires de nombres arabes qui composent ces premiers items, ne coïncident pas. Ce type d'interférence peut générer plus d'erreurs (Rubinstein et al., 2002). Malgré le nombre important de bonnes réponses à ces items, les indices de discrimination demeurent très satisfaisants selon la classification d'Ebel (1965). Nous avons donc décidé de les conserver dans l'objectif de pouvoir différencier plus finement les sujets faibles. Il serait cependant intéressant de valider ce résultat auprès d'une plus grande population.

De la même manière que pour les catégories 1, 2 et 3, il apparaît nécessaire d'établir un nouvel ordre de présentation des items de la catégorie 4 pour garantir une progression de difficulté croissante.

2.2.3. Items de la catégorie 5 (fractions et nombres décimaux)

En ce qui concerne les items de la catégorie 5 (49 à 60), les résultats du Tableau 8 montrent que les élèves de CP, CE1 et CE2 les réussissent moins bien que les élèves de CM1 et CM2. Cette observation suggère que les élèves de ces trois niveaux inférieurs ne possèdent pas tous les compétences requises pour réussir la dernière catégorie d'items, qui consiste à comparer des fractions et des nombres décimaux, conformément à leur niveau scolaire (Ministère de l'Education Nationale et de la Jeunesse, 2020). La différence est particulièrement marquée entre les élèves de CP et CM2, où l'on observe de manière respective des effets planchers (0% de réussite) et des effets plafonds (100% de réussite) pour les items 50 à 60. Cela explique pourquoi ces items ne discriminent pas suffisamment les compétences entre les élèves de CP d'une part, puis entre les élèves de CM2 d'autre part.

Deux items se démarquent par des taux de réussite faibles, tous niveaux confondus.

L'item 54 (1,75 vs 6/3) possède un taux de réussite de 29%. Ce résultat n'est pas surprenant car la comparaison entre un nombre décimal et une fraction n'est pas un concept mathématique attendu au cours du cursus élémentaire. Cependant, les items 52 et 60 évaluant ce type de comparaison ont montré un niveau de difficulté moyen. Un effet de distance peut expliquer le taux d'erreurs plus important pour cet item, étant donné que le ratio entre 1,75 et 6/3, égal à 0,87, est plutôt faible. Il nous semble donc intéressant de conserver cet item afin de permettre la comparaison des performances de l'enfant entre les items 54, 52 et 60, dont les paires varient en fonction de la taille du ratio.

L'item 49 (1/4 vs 1/7), quant à lui, se distingue par une difficulté significativement élevée, avec un taux de réussite inférieur à 10%. La complexité de cet item peut être attribuée à plusieurs facteurs. D'une part, les fractions 1/4 et 1/7 sont présentées dans l'ordre inverse de la ligne numérique mentale, ce qui est source d'erreurs (effet SNARC). D'autre part, ces fractions ne possèdent pas le même dénominateur. Or, la comparaison de fractions avec un dénominateur différent ne fait pas partie du programme scolaire élémentaire. L'item maintient par ailleurs un bon indice de discrimination pour les élèves de CM1 et CM2.

Nous faisons finalement le choix de conserver les items de la cinquième catégorie, dans la perspective de les réserver aux niveaux scolaires plus avancés tels que le CM1 et le CM2. Cette décision vise à garantir leur pertinence pour ces niveaux ultérieurs et à éviter une évaluation inappropriée des comparaisons de fractions et de nombres décimaux pour les élèves de CP, CE1 et CE2. Cependant, une vérification des résultats auprès d'un échantillon plus large est nécessaire pour déterminer si l'item 49 doit être conservé ou modifié.

Face à ces résultats, une approche prudente consisterait à conserver tous les items de notre épreuve, afin d'assurer des niveaux de difficulté variés et un niveau suffisant d'items. Cependant, quelques perspectives doivent être envisagées.

3. Limites et perspectives

3.1. Contenu de l'épreuve

Malgré la corrélation observée entre l'organisation des catégories d'items et le niveau développemental ou scolaire des élèves dans les résultats principaux, l'analyse par item a révélé que ces derniers ne suivent pas un ordre croissant de difficulté. Les indices de difficulté par item reportés dans le tableau en annexe 3, montrent en effet que certains items sont mal ordonnés. Afin de maintenir la cohérence des catégories de notre épreuve, nous avons procédé à un autre classement, qui prend en considération à la fois les catégories d'items et la progression ascendante de la difficulté. Ce classement, disponible en annexe 4, permet de présenter les items les plus faciles en premier et les items les plus compliqués en dernier, dans chaque catégorie.

Par ailleurs, il a été démontré que certaines catégories d'items ne sont pas nécessairement pertinentes selon les niveaux scolaires. Comme cela avait été envisagé par Cousin (2021), l'épreuve devra inclure à terme des critères de début et de fin en fonction du niveau développemental ou scolaire de l'enfant, afin de permettre l'élimination des items trop faciles ou trop difficiles pour un sujet. Autrement dit, l'enfant ne commencera pas à la catégorie 1 contenant des « petits nombres » si la comparaison de ces derniers est censée être acquise en regard de son niveau scolaire. Néanmoins, s'il ne parvient pas à comparer les items de la catégorie correspondant à son niveau, un retour en arrière pourra être possible pour identifier les comparaisons de nombres qu'il est capable de réaliser. Des sauts de catégories et des critères d'arrêt, initialement pensés par Cousin (2021), peuvent également être envisagés si, de manière respective, l'enfant répond correctement ou échoue à six items consécutifs. Ce critère d'arrêt permettrait d'éviter de mettre l'enfant en trop grande difficulté.

3.2. Population

Sur un total de 80 sujets recrutés, 79 enfants ont participé à notre étude, ce qui représente un nombre restreint. Répartis selon les niveaux scolaires, les groupes ne sont pas représentés de manière équitable et contiennent au maximum 23 sujets. Bien que nous ayons pu obtenir quelques tendances statistiques à partir de cet échantillon, la représentativité statistique est donc limitée.

Aucun critère d'exclusion n'a été établi dans le cadre de notre étude. Ce choix découle principalement de notre volonté d'explorer et de comprendre la variabilité naturelle des performances des participants, en veillant à maximiser la diversité et la représentativité de notre échantillon. Cependant, nous avons jugé pertinent de relever les suivis orthophoniques éventuels des participants, afin d'évaluer leur impact sur leurs résultats à l'épreuve. Parmi les sujets, sept enfants ont bénéficié d'un suivi orthophonique pour des troubles autres que ceux liés aux apprentissages des mathématiques. Une observation révèle que pour six d'entre eux, les scores moyens obtenus sont légèrement inférieurs à la moyenne des enfants du même niveau scolaire. Cette constatation n'exclut pas la possibilité d'une influence de ces troubles sur les performances des enfants en comparaison de nombres arabes.

Par ailleurs, nous nous sommes intéressés aux enfants scolarisés du CP au CM2 pour la phase de pré-test, principalement pour des raisons d'organisation pratique des passations. Cependant, les résultats de notre étude ont mis en évidence des effets plafonds dès le CP pour les premières catégories d'items, ainsi que des effets plafonds pour les élèves de CM2 à la dernière catégorie d'items, signifiant leur niveau de maîtrise suffisant pour réussir ces items.

En considérant l'ensemble de ces éléments, il serait nécessaire de poursuivre le recrutement afin d'obtenir des groupes plus homogènes et des résultats plus significatifs. Les données collectées seraient en effet plus représentatives de la population cible, permettant alors une interprétation plus fiable et généralisable des résultats. Étendre l'échantillon en incluant des élèves de Grande Section de Maternelle et de 6ème serait également intéressant pour couvrir de plus grandes classes d'âge et donc une plus grande diversité de compétences.

3.3. Passations

Dans le cadre de la création d'une nouvelle batterie d'évaluation diagnostique, les participants ont été soumis à sept tests différents, incluant notre épreuve de comparaison de nombres arabes. Sur une trentaine de minutes de passation, 147 secondes, soit 2 minutes et 27 secondes, ont été nécessaires en moyenne pour terminer l'épreuve de comparaison de nombres. Par sa courte durée, l'épreuve garantit une évaluation rapide de cette compétence. Cette constatation confirme la faisabilité de l'épreuve et est d'autant plus intéressante dans le contexte de la conservation des items, car elle suggère que le nombre total de 60 items n'allonge pas excessivement la durée de l'épreuve. En raison de sa rapidité et de sa simplicité, cette épreuve pourrait également servir à mettre l'enfant en confiance dès le début de l'évaluation.

Pour garantir des conditions de passation représentatives de la pratique clinique, nous avons opté pour des séances de passation uniques, organisées en situation duelle, dans un environnement calme. Cette approche visait à minimiser les distractions potentielles et à permettre aux participants de se concentrer pleinement sur la tâche à accomplir, contribuant ainsi à des conditions de passation optimales pour notre épreuve. De plus, considérant que dans la pratique clinique, les enfants ne débutent pas toujours par la même épreuve en fonction des objectifs d'évaluation, toutes les épreuves ont été administrées dans un ordre aléatoire. Cependant, bien que cette randomisation puisse refléter la réalité clinique, elle peut également introduire d'autres facteurs susceptibles d'influencer les performances des participants. Par exemple, si une épreuve plus difficile est présentée avant notre épreuve de comparaison de nombres arabes, cela peut entraîner une fatigue ou une diminution de la confiance en soi, ce qui peut à son tour affecter les performances à notre épreuve. Par conséquent, nous ne pouvons exclure l'impact de ces facteurs sur les résultats obtenus.

L'épreuve de comparaison de nombres arabes a été présentée sous format « papier-crayon ».

La consigne, lue pour chaque participant, était inscrite directement sur le protocole, au recto de la feuille. L'analyse qualitative a démontré que l'instruction avait été comprise par tous les participants : ces derniers savaient qu'ils devaient identifier le nombre arabe le plus grand dans chaque paire. Cependant, au lieu de les barrer comme indiqué, quatre enfants ont entouré les items. Cette action peut être attribuée à un manque d'attention plutôt qu'à un manque de compréhension. En-dessous de la consigne, les deux items d'exemple étaient présentés. D'après l'analyse qualitative, aucun enfant n'a rencontré de problème avec ces items.

Au verso du protocole, cinq colonnes étaient présentées, chacune contenant douze rectangles représentant respectivement les cinq catégories et les douze paires de nombres à comparer. Chaque paire de nombres était écrite en noir sur fond blanc, utilisant la même police et la même taille de caractères. Alors que cette uniformité visuelle est optimale pour réduire tout biais éventuel associé aux caractéristiques visuo-spatiales des symboles numériques (Cohen Kadosh & Henik, 2006;

Schwarz & Ischebeck, 2003), le format papier peut présenter quelques limites. Tout d'abord, la présentation de toutes les paires de nombres sur une seule feuille peut décourager l'enfant en lui donnant une vue d'ensemble de tous les items à comparer. Un enfant de CP peut par exemple ressentir de l'appréhension en apercevant dès le début de l'épreuve des nombres qu'il ne connaît pas, notamment des fractions et des nombres décimaux. Par ailleurs, l'analyse qualitative des passations a mis en évidence une variabilité dans la modalité des réponses, en lien avec le format papier : certains ont barré d'un seul trait, d'autres ont réalisé une croix. Enfin, nos capacités de notation semblent limitées par ce support. Alors que nous avons facilement accès au temps de réponse total et au temps de réponse par catégorie, nous n'avons pas pu chronométrer de façon exacte le temps de réponse par item. Cependant, ce dernier permet de mieux cibler les difficultés en fonction des comparaisons erronées, au même titre que les scores par item. Ces mesures sont par ailleurs essentielles pour analyser la capacité du sujet à comparer deux nombres arabes, et obtenir des informations sur le niveau d'automatisation du processus de traitement des nombres et sur la capacité d'accès au sens du nombre par le code arabe.

Comme cela avait été présenté par Cousin (2021), une informatisation de l'épreuve pourra être envisagée pour la suite. Elle permettra de pallier les limites identifiées en présentant un item à la fois, en calculant automatiquement les scores et les temps de réponse de manière rapide et exacte, et en ne proposant qu'une modalité de réponse. Cela impliquera une adaptation de la consigne au format informatisé et donc une détermination du type de réponse motrice disponible (cliquer avec la souris sur la bonne réponse ou appuyer sur une touche du clavier selon la position de la bonne réponse sur l'écran).

Conclusion

La capacité à comparer des nombres arabes émerge précocement dans le développement de la cognition numérique. Son amélioration est progressive, influencée par les apprentissages scolaires et l'évolution des mécanismes cognitifs. Pourtant, certains enfants ne parviennent pas à développer une représentation mentale précise des nombres et éprouvent alors des difficultés à traiter les relations de magnitude entre les nombres symboliques. Cette observation a été constatée chez les enfants présentant un TAM (De Smedt & Gilmore, 2011; Desoete et al., 2012). Pour aider au diagnostic précoce de ce trouble, Cousin (2021) a développé un test de comparaison de nombres arabes, à partir du courant théorique actuel et d'une analyse critique des protocoles existants. Chronométré, ce test comporte soixante items à comparer, répartis en cinq catégories de complexité croissante.

L'objectif de cette présente étude était de réaliser le pré-test de cette épreuve, afin d'évaluer sa pertinence et la qualité de son contenu auprès de sujets témoins. Au total, 79 enfants, répartis en cinq niveaux scolaires (du CP au CM2), se sont vu administrer l'épreuve de comparaison de nombres arabes.

Les résultats obtenus par l'ensemble des sujets témoins ont permis de vérifier que les performances en comparaison de nombres arabes évoluent en fonction du niveau scolaire et de la complexité des catégories d'items. Cela suggère la pertinence d'une épreuve évaluant cet aspect. Cependant, des items se sont avérés inadaptés à certains niveaux scolaires en raison de leur niveau de difficulté. Il semble donc pertinent de considérer la proposition de Cousin (2021) concernant une cotation adaptée au niveau du sujet. Les items ont été ordonnés selon une progression de difficulté croissante, en vue de cette possibilité.

Qualitativement, l'épreuve s'est révélée facilement réalisable et accessible, grâce à sa durée d'application rapide ainsi qu'à la compréhension aisée de la consigne. Son format pourrait cependant être amélioré par une informatisation, offrant une présentation optimale des items au sujet ainsi qu'une notation rapide et précise pour l'examinateur.

Il apparaît désormais essentiel de poursuivre cette étude afin de compléter l'échantillon de notre population et de l'élargir à davantage de classes. Cette approche permettrait d'observer une plus grande diversité de compétences et de parvenir à une décision définitive concernant la conservation de certains items. Pour améliorer davantage l'épreuve, il serait également intéressant d'apporter les ajustements recommandés, notamment en ce qui concerne l'ordre de présentation des items, l'utilisation d'un support informatisé et les critères de cotation.

Une telle épreuve pourrait permettre aux orthophonistes, à terme, de disposer d'un outil évaluant la précision et la rapidité d'accès à la magnitude à partir du code arabe, permettant ainsi d'étayer leur diagnostic et d'adapter au mieux leur prise en soin.

Bibliographie

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders, 5th Edition : DSM-5 (5e éd.). *American Psychiatric Publishing*.

Ashkenazi, S., Mark-Zigdon, N., & Henik, A. (2009). Numerical distance effect in developmental dyscalculia. *Cognitive Development, 24*(4), 387–400.

Borst, G., Aïte, A., & Houdé, O. (2015). Inhibition of misleading heuristics as a core mechanism for typical cognitive development: Evidence from behavioral and brain-imaging studies. *Developmental Medicine Child Neurology, 57*(2), 21-25.

Brannon, E. (2006). The representation of numerical magnitude. Curent opinion in *Neurobiology, 16*(2), 222-229.

Butterworth, B. (2005). Developmental dyscalculia. Dans J.I.D. Campbell (dir.), *Handbook of mathematical cognition* (p. 455–467). Psychology Press.

Chartier, P., & Vrignaud, P. (2018). Les méthodologies d'évaluation. Dans P. Chartier (dir.), *L'orientation scolaire et professionnelle: Pratiques d'évaluation* (p. 23-59). Mardaga.

Chazoule, G., & Thevenot, C. (2018). Apprendre à manipuler les symboles numériques. Dans L. Ferrand, B. Lété, C. Thevenot (dir.), *Psychologie cognitive des apprentissages scolaires : Apprendre à lire, écrire, compter* (p. 265-278). Dunod.

Cohen Kadosh, R., & Henik, A. (2006). A common representation for semantic and physical properties : a cognitive-anatomical approach. *Experimental psychology, 53*(2), 87–94.

Cohen Kadosh, R., Tzelgov, J., & Henik, A. (2008). A synesthetic walk on the mental number line : The size effect. *Cognition, 106*(1), 548-557.

Cousin, E. (2021). *Proposition d'évaluation de la comparaison symbolique dans un projet de création de bilan orthophonique de la cognition mathématique : création de l'épreuve de comparaison de nombres arabes* [mémoire de master, Université de Lille]. Pépite. https://pepite-depot.univ-lille.fr/RESTREINT/Mem_Ortho/2021/LILU_SMOR_2021_026.pdf

Davidson, M.-C., Amso, D., Anderson, L.-C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037–2078.

De Smedt, B., & Gilmore, C.-K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology, 108,* 278-292.

De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends Neurosci. Educ. 2*, 48–55.

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of experimental child psychology, 103*(4), 469–479.

Dehaene S. (1992). Varieties of numerical abilities. *Cognition, 44*(1-2), 1–42.

Dehaene S. (2010). *La bosse des maths*. Odile Jacob.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology*, *122*(3), 371–396.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3-6), 487–506.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science (New York, N.Y.), 284*(5416), 970–974.

Demeuse, M., & Henry, G. (2004). L'analyse classique d'items. Dans M. Demeuse (dir.), *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation* (p. 173-186). Les éditions de l'Université de Liège.

Dempster, F.-N. (1992). The rise and fall of the inhibitory mechanism: toward a unify theory of cognitive development and aging. *Developmental Review, 12*, 45-75.

Desoete, A., Ceulemans, A., De Weerdt, F., & Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten ? Findings from a longitudinal study. *The British journal of educational psychology, 82*(Pt 1), 64–81.

Durand, M., Hulme, C., Larkin, R., & Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7- to 10-year-olds. *Journal of Experimental Child Psychology, 91*(2), 113–136.

Ebel, R.-L. (1965). *Measuring educational achievement*. Prentice-Hall.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, *8*(7), 307–314.

Gilmore, C.-K., McCarthy, S. E., & Spelke, E.-S. (2007). Symbolic arithmetic knowledge without instruction. *Nature*, *447*(7144), 589–591.

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of experimental child psychology*, *76*(2), 104–122.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental psychology*, *44*(5), 1457–1465.

Hawes, Z., Nosworthy, N., Archibald, L.-M.-D., & Ansari, D. (2019). Kindergarten children's symbolic number comparison skills relates to 1st grade mathematics achievement : Evidence from a two-minute paper-and-pencil test. *Learning And Instruction*, *59*, 21-33.

Holloway, I.-D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. *Journal of experimental child psychology*, *103*(1), 17–29.

Houdé, O. (2018). *Le raisonnement*. Presses Universitaires de France.

Iuculano, T., Tang, J., Hall, C.-W.-B., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science, 11*, 669-680.

Jordan, N.-C., Kaplan, D., Ramineni, C., & Locuniak, M.-N. (2009). Early math matters : Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*(3), 850-867.

Lafay, A., & Cattini, J. (2018). Analyse psychométrique des outils d'évaluation mathématique utilisés auprès des enfants francophones. *Canadian Journal of Speech-Language Pathology and Audiology*, *42*(2),127-144.

Lafay, A., St-Pierre M.-C., & Macoir, J. (2013). Développement des systèmes numériques non symboliques et prédicteurs de réussite mathématique. *Glossa, 112,* 1-17.

Lafay, A., St-Pierre M.-C., & Macoir, J. (2014). L'évaluation des habiletés mathématiques de l'enfant : inventaire critique des outils disponibles. *Glossa, 116,* 33-58.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities : A study of 8-9-year-old students. *Cognition, 93*, 99-125.

Landerl, K., Fussenegger, B., Moll, K., & Willburger, E. (2009). Dyslexia and dyscalculia: Two learning disorders with different cognitive profiles. *Journal of Experimental Child Psychology, 103*, 309-324.

Landerl, K., & Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *Journal of Experimental Child Psychology, 103*, 546-565

Laski, E.-V., & Siegler, R.-S. (2007). Is 27 a big number ? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child development, 78*(6), 1723–43.

Laveault, D., & Grégoire, J. (2023). *Introduction aux théories de tests en psychologie et sciences de l'éducation* (4e éd.). De Boeck Supérieur.

Lipton, J.-S., & Spelke, E.-S. (2003). Origins of Number Sense. *Psychological Science, 14,* 396 - 401.

Lipton, J.-S., & Spelke, E.-S. (2005). Preschool Children's Mapping of Number Words to Nonsymbolic Numerosities. *Child Development, 76*(5), 978–988.

Ministère de l'Education Nationale et de la Jeunesse. (2020). *Bulletin officiel n°31 du 30 juillet 2020.* https://www.education.gouv.fr/sites/default/files/imported_files/documents/BO_MENJS_31_1313155.pdf

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance - A longitudinal study on numerical development. *Research in Developmental Disabilities, 32*(5), 1837-1851.

Mundy, E., & Gilmore, C.-K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology*, *103*(4), 490–502.

Mussolin, C., Mejias, S., & Noël, M.-P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition*, *115*(1), 10–25.

Nosworthy, N., Bugden, S., Archibald, L.-M.-D., Evans, B.-J., & Ansari, D. (2013). A Two-Minute Paper-and-Pencil Test of Symbolic and Nonsymbolic Numerical Magnitude Processing Explains Variability in Primary School Children's Arithmetic Competence. *Plos One*, *8*(7), e67918.

Obersteiner, A., Van Dooren, W., Van Hoof, J., & Verschaffel, L. (2013). The natural number bias and magnitude representation in fraction comparison by expert mathematicians. *Learning and Instruction, 28*, 64-72.

Piazza, M., Facoetti, A., Trussardi, A.-N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33–41.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, *53*(2), 293–305.

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron, 41*(6), 983–993.

Price, G.-R., Holloway, I., Rasanen, P., Vesterinen, M., & Ansari, D. (2007). Impaired parietal magnitude processing in developmental dyscalculia. *Current Biology, 17*(24), 1042-1043.

Roell, M., Viarouge, A., Houdé, O., & Borst, G. (2017). Inhibitory control and decimal number comparison in school-aged children. *Plos One, 12*(11), e0188276.

Roell, M., Viarouge, A., Houdé, O., & Borst, G. (2019). Inhibition of the whole number bias in decimal number comparison: A developmental negative priming study. *Journal of Experimental Child Psychology*, *177*, 240-247.

Rousselle, L., & Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic versus non-symbolic number magnitude processing. *Cognition*, *102*, 361-395.

Rousselle, L., Palmers, E., & Noël, M.-P. (2004). Magnitude comparison in preschoolers: what counts? Influence of perceptual variables. *Journal of experimental child psychology*, *87*(1), 57–84.

Rubinsten, O., & Henik, A. (2005). Automatic activation of internal magnitudes: a study of developmental dyscalculia. *Neuropsychology*, *19*(5), 641–648.

Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *Journal of experimental child psychology*, *81*(1), 74–92.

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental science, 20*(3).

Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number-size interference in comparative judgments of numerals. *Journal of Experimental Psychology: Human Perception & Performance, 29*, 507–522.

Schwenk, C., Sasanguie, D., Kuhn, J.-T., Kempe, S., Doebler, P., & Holling, H. (2017). (Non-)symbolic magnitude processing in children with mathematical difficulties: A meta-analysis. *Research in Developmental Disabilities, 64*, 152-167.

Siegler, R.-S., & Pyke, A.-A. (2013). Developmental and individual differences in understanding of fractions. *Developmental psychology, 49*(10), 1994–2004.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(1), 166–179.

Vanbinst, K., Ansari, D., Ghesquière, P., & De Smedt, B. (2016). Symbolic numerical magnitude processing is as important to arithmetic as phonological awareness is to reading. *Plos One, 11*(3).

Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2012). Naturally biased? In search for reaction time evidence for a natural number bias in adults. *The Journal of Mathematical Behavior, 31*(3), 344–355.

Von Aster, M., & Shalev, R.-S. (2007). Number development and developmental dyscalculia. *Developemental Medicine & Child Neurology, 49*, 868–873.

Wahl, G., & Wahl, M. (2020). Chapitre III. La dyscalculie. Dans G. Wahl (dir.), *Les enfants DYS* (p. 34-48). Presses Universitaires de France.

White, S.-L., Szűcs, D., & Soltész, F. (2012). Symbolic number: the integration of magnitude and spatial representations in children aged 6 to 8 years. *Frontiers in psychology*, *2*, 392.

Wilson, A.-J., & Dehaene, S. (2007). Number Sense and Developmental dyscalculia. Dans D. Coch, E. Dawson, et X. Fischer (dir.), *Human Behavior learning, and the developping brain: atypical development* (p. 212-238). The Guilford Press.

Xu, F., & Spelke, E.-S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*, B1-B11.

Liste des annexes