

ARTICULER DES MÉTHODES DE RECHERCHE EN DIDACTIQUE DES MATHÉMATIQUES ET EN SCIENCES COGNITIVES : UN EXEMPLE DE CONCEPTION ET VALIDATION D'UN TEST

GARDES* MARIE-LINE, CROSET** MARIE-CAROLINE, PRADO*** JÉRÔME ET SAMSON****

ADELINE

Résumé | Au sein d'études interventionnelles se pose la question du choix des tests pour évaluer les connaissances et les progrès des élèves avant et après la séquence d'enseignement. L'analyse didactique des tests existants nous a conduits à concevoir nous-mêmes un test, afin qu'ils puissent répondre plus précisément à nos questions de recherche. Nous souhaitons mettre en évidence l'apport de nos différents regards (statistique, didactique et cognitif) dans la conception et la validation d'un test pour évaluer les connaissances des élèves sur la construction du nombre à l'école maternelle.

Mots-clés : apprentissage du nombre, évaluation de connaissances, école maternelle, cognition numérique, didactique des mathématiques

Abstract | In intervention studies, the question arises of the choice of tests for assessing students' knowledge and progress before and after the teaching intervention. The didactic analysis of existing tests led us to design the tests ourselves, so that they could respond more precisely to our research questions. Our aim is to highlight the contribution of our different perspectives (statistical, didactic and cognitive) in the design and validation of a test to assess students' knowledge of the concept of number in pre-school.

Keywords: Number learning, knowledge assessment, preschool, cognition, mathematics didactics

I. CONTEXTE

Dans un contexte où les performances mathématiques des élèves français dans les évaluations internationales telles que TIMSS et PISA suscitent des préoccupations, le gouvernement français a encouragé le développement d'études interventionnelles¹. L'une des nécessités de ces études est l'élaboration d'outils d'évaluation des connaissances adaptés à différents âges, permettant d'analyser de manière satisfaisante le développement des connaissances des enfants. Ces études, menées en classe, impliquent souvent la collaboration de chercheurs issus de divers domaines, par exemple la psychologie cognitive et la didactique des mathématiques. Nous pensons que cette interdisciplinarité enrichit les études par la diversité des perspectives et des méthodologies mais impose également un défi : la création de tests de connaissances capables de satisfaire les exigences de l'ensemble des chercheurs concernés.

Dans cet article, nous proposons de décrire un travail impliquant psychologues et didacticiens autour de la conception et de l'évaluation d'un test de connaissances du concept de nombre destiné à des enfants de 5-6 ans. Nous commençons par présenter un modèle didactique des connaissances sur le nombre sur lequel s'est basée la conception du test. Nous décrivons ensuite la méthode utilisée pour

* Haute École Pédagogique Vaud, UER MS, Lausanne – Suisse – marie-line.gardes@hepl.ch

** Université Grenoble Alpes, CNRS, LIG, Grenoble – France – marie-caroline.croset@univ-grenoble-alpes.fr

*** Centre de Recherche en Neurosciences de Lyon (CRNL), INSERM U1028 - CNRS UMR5292, Université de Lyon – France – jerome.prado@univ-lyon1.fr

**** Université Grenoble Alpes, CNRS, LJK, Grenoble – France – adeline.leclercq-samson@univ-grenoble-alpes.fr

¹ Une étude interventionnelle vise à évaluer les effets d'un enseignement spécifique sur les apprentissages des élèves.

concevoir chaque item du test, de manière à contraindre les techniques des élèves et à identifier précisément les connaissances mobilisées. Enfin, nous présentons les analyses statistiques menées pour valider ce test. En conclusion, nous discutons de l'intérêt de combiner analyses quantitatives et qualitatives pour offrir une image précise et fidèle des connaissances des élèves.

II. MODÈLE DIDACTIQUE : UNE CARTE DES CONNAISSANCES SUR LE NOMBRE

Chevallard introduit la notion de rapport personnel qu'un élève entretient avec un objet de savoir comme un concept unificateur des aspects fragmentaires sous lesquels on décrit communément la connaissance (Chevallard, 1989). Ainsi regroupe-t-il les notions de savoir-faire, conceptions et savoir au sein de la notion du rapport personnel à un objet de savoir. Bosch et Chevallard (1999) proposent un quadruplet praxéologique pour décrire les connaissances des élèves attendues par une institution : un type de tâches accompli par des techniques qui se justifient par des technologies au sein d'une théorie. En 2019, nous avons décrit le rapport personnel que pouvait entretenir un élève sur le concept de nombre chez un élève de maternelle en France (Croset et Gardes, 2019, 2020). 13 types de tâches ont été identifiés et catégorisés selon l'usage du nombre : cardinal, ordinal et, enfin, un usage du nombre en dehors de toute référence à une grandeur physique (un usage décontextualisé). Pour chaque usage, nous avons identifié cinq grands types de tâches, à savoir le codage, le décodage, l'association d'un nombre à un autre nombre, la comparaison et l'anticipation du résultat d'une action. À partir de travaux issus de la psychologie cognitive et de la didactique des mathématiques, nous proposons de préciser ci-dessous les techniques principales associées aux différents types de tâches (Annexe 1). Ponctuellement, nous mentionnons aussi des éléments technologiques justifiant certaines techniques.

Coder une quantité ou une position. Ce type de tâches consiste à déterminer le cardinal (le nombre d'éléments) d'une collection ou la position d'un élément dans une liste ordonnée d'objets. Par exemple, il peut être demandé à un enfant « combien y a-t-il de jetons devant toi ? » ou « à quelle position se situe la perle verte dans ce collier ? ». La réponse peut être donnée par un mot-nombre ou un code arabe. Différentes techniques permettent de coder tels que le *subitizing* (capacité de reconnaître immédiatement et précisément le cardinal de petites quantités (Clements, 1999), le comptage (de un en un ou, plus généralement, de n en n) consistant à mettre en correspondance le mot-nombre de la comptine conventionnelle et un élément de la collection. Le comptage mobiliserait différents éléments technologiques, notamment les cinq principes de Gelman (2009), la notion d'unité et d'itération.

Décoder un nombre par une quantité ou une position. Ce type de tâches consiste à construire une collection d'un cardinal donné (par un mot-nombre ou par une écriture chiffrée) (Wynn, 1992). Les consignes « Peux-tu me donner 5 jetons ? » ou « Peux-tu désigner la 5e perle ? » en sont des exemples. Ce type de tâches est la réciproque de celui de codage : on part d'un code symbolique (mot-nombre ou écriture chiffrée) et on lui associe une quantité. Différentes techniques similaires aux précédentes permettent de décoder : le *subitizing* ou le comptage. Elles impliquent le stockage du nombre cible en mémoire de travail, le processus d'énumération (Briand, 1999) et de contrôle pour s'arrêter lorsque le nombre cible est atteint.

Associer des quantités ou des positions. Ce type de tâches consiste à construire une collection de même cardinal qu'une autre collection témoin, mais aussi à désigner une position équivalente à une position témoin dans un ensemble ordonné (Cariat et Margolinas, 2015). Par exemple, il pourrait être demandé à un enfant « Voici des stylos, peux-tu prendre autant de bouchons que de stylos ? ». Contrairement aux deux types de tâches précédents, le code symbolique (mot-nombre ou écriture chiffrée) n'est pas nécessairement présent dans une tâche d'association. Nous avons identifié trois

techniques pour ce type de tâches : la perception, la correspondance terme à terme et le codage/décodage. La perception permet de conserver la quantité de la collection témoin puis de construire la collection de même quantité. La correspondance terme à terme consiste à associer un élément de la collection à construire à un élément de la collection témoin. Selon le contexte matériel, cela peut se ramener à relier d'un trait chaque élément, les pointer simultanément ou encore déposer un élément face à chaque élément témoin. Enfin, le codage/décodage consiste à coder puis décoder, c'est-à-dire, déterminer le cardinal de la collection témoin ou la position (par une technique de codage : *subitizing*, comptage de un en un, etc.) puis construire la collection de ce cardinal ou la position équivalente (par une des techniques de décodage). Cette technique est la seule technique utilisant le nombre pour ce type de tâches.

Comparer des quantités ou des positions. Ce type de tâches consiste, pour deux collections données (ou pour deux positions dans des listes ordonnées), à désigner la collection (ou la position) qui a le plus ou le moins d'éléments (ou la plus proche ou la plus éloignée de l'origine). Par exemple, une consigne pour ce type de tâches pourrait être « Voici des jetons rouges. Voici des jetons bleus. Peux-tu me dire s'il y a plus de jetons rouges que de jetons bleus ? ». Trois techniques sont mobilisables : l'estimation, la correspondance terme à terme et le codage. L'estimation est une procédure qui serait innée et reposerait sur une représentation de la magnitude des nombres (Dehaene, 2018). Elle est fondamentalement approximative. La correspondance terme à terme consiste à associer un élément d'une collection à un élément de l'autre collection. À la fin de cette correspondance terme à terme, il faut conclure que la quantité la plus grande est celle dans laquelle il reste des éléments non appariés. Le codage consiste à coder la quantité de chaque collection (par une des techniques de codage citées précédemment) et comparer les codes symboliques obtenus. En plus des éléments technologiques du codage, il faut donc ajouter la connaissance de l'ordre des nombres.

Opérer sur des quantités ou des positions. Ce type de tâches consiste à « anticiper » un résultat après une action (ajouter, enlever, partager, *etc.*) ou une position après une action (avancer, reculer, *etc.*). Par exemple, une consigne pour ce type de tâches pourrait être « Il y a 3 jetons dans la boîte (opaque). J'en ajoute 2. Combien y a-t-il de jetons dans la boîte ? » Pour obtenir le résultat, plusieurs techniques sont disponibles : le recomptage (compter à partir du début), le calcul, le surcomptage ou le sous-comptage (compter en avant ou en arrière à partir d'un nombre cardinal/position donné) et la connaissance des faits numériques.

À ces 10 types de tâches, s'ajoutent 3 types de tâches dans un usage décontextualisé : associer le code verbal au code arabe et réciproquement (c'est-à-dire transcoder), comparer des nombres (par exemple dans un jeu de batailles où les cartes ne comportent que des écritures chiffrées) et calculer sur des nombres (par exemple, trouver la carte correspondante à la somme de deux cartes).

III. CONCEPTION ET PRÉSENTATION DU TEST

1. Conception du test

Notre objectif était de proposer des tâches aux élèves qui imposent de ne mobiliser qu'une et une seule technique afin de pouvoir diagnostiquer les technologies mobilisées par l'élève. Pour cela, nous avons identifié la portée de chaque technique :

Une technique — une « manière de faire » — ne réussit que sur une partie $P(t)$ des tâches du type T auquel elle est relative, partie qu'on nomme portée de la technique : elle tend à échouer sur $T \setminus P(t)$, de sorte qu'on peut dire que « l'on ne sait pas, en général, accomplir les tâches du type T » (Chevallard, 1998, p. 225).

À la manière de T4tel (Chaachoua et Bessot, 2016), nous avons décrit ces portées comme un ensemble de valeurs prises par des variables. Par exemple, pour le type de tâches « Comparer des quantités », six tâches sont proposées : deux pour évaluer la technique comptage, deux pour évaluer la technique de correspondance terme à terme et deux pour évaluer la technique d'estimation. Deux variables didactiques principales sont en jeu dans ce type de tâches : l'écart entre les deux quantités et l'accès simultané des collections. La comparaison de deux quantités par estimation est facilitée lorsque ces quantités sont de cardinaux très différents. La loi de Weber explique cet effet de distance : « l'imprécision de l'estimation, mesurée par son écart type, croît linéairement avec le nombre estimé, selon une constante de proportionnalité appelée fraction de Weber » (Dehaene, 2008, p. 280). La portée de la technique d'estimation est donnée par la fraction de Weber, qui dépend de l'âge (Halberda et Feigenson, 2008 ; Piazza et al., 2010) : pour les enfants de 5 ans, cette fraction varie de 0,23 à 0,34. La portée de la technique de mise en correspondance est que les éléments des deux collections doivent être visibles simultanément ou déplaçables.

Dans ce travail d'identification des tâches mobilisant telle technique, il faut aussi décrire les valeurs de variables qui empêcheront la mobilisation d'une autre technique. Par exemple, pour empêcher l'estimation, il faut que l'écart entre les deux quantités à comparer ne soit pas important. Une disposition rectiligne des objets peut aussi engager à mobiliser la correspondance terme à terme et éviter le comptage. Par ailleurs, comme l'âge des enfants impose de proposer des collections de quantité raisonnable, les objets de la collection doivent être de taille différente afin que la surface occupée par les objets ne puisse pas permettre de répondre. La technique de codage, quant à elle, a une portée plus importante et devient nécessaire dès lors que les deux autres techniques sont empêchées. Les différents items de comparaison de quantités sont détaillés dans le tableau 1 où la tâche consistait à désigner la collection ayant le plus d'étoiles.

Tableau 1 – Exemples de variables pour le type de tâches « Comparer des quantités »

Technique	Portée de la technique				Exemple de tâche
	Variable Taille étoile	Variable Simultanéité	Variable Organisation	Variable Écart	
Coder par comptage	de taille différente ou d'espace différent entre les étoiles	collections séparées, par une feuille blanche, non visibles simultanément	désorganisée	écart faible	
Correspondre terme à terme		collections visibles simultanément	rectiligne	écart faible	
Estimer		collections visibles simultanément	désorganisée	écart important	

2. Présentation du test ENUDIT 2

Le test est structuré en trois parties, selon les trois usages du nombre : cardinal (partie A), ordinal (partie B) et décontextualisé (partie C). Le tableau 2 présente les tâches des parties A et B, le tableau 3 présente celles de la partie C. Chaque partie est découpée selon les cinq types de tâches décrits précédemment. Les parties A et B comportent ainsi les types de tâches suivants : coder (A1, B1), décoder (A2, B2), comparer (A3, B3), associer (A4, B4) et anticiper (A5, B5). La partie C a trois types de tâches : transcoder (C3), comparer (C4) et anticiper (C5). Un type de tâches a été ajouté dans cette partie : réciter la comptine numérique (C0). Pour chaque type de tâches, plusieurs tâches sont proposées afin d'évaluer les différentes techniques identifiées. Pour chaque type de tâches, les éléments sont présentés par ordre décroissant de difficulté, selon la technique mobilisée². Si l'enfant réussit les items les plus difficiles, il n'est pas exposé aux items considérés comme plus faciles : un présupposé du test est qu'il serait capable de les réussir puisqu'il a réussi des tâches similaires plus complexes. Les mêmes valeurs ont été choisies pour les deux parties A et B (par exemple, six étoiles dans la tâche A1 et le sixième étage dans la tâche B2). L'idée est de pouvoir comparer si, pour une même valeur numérique, les enfants réussissent mieux dans les tâches cardinales que dans les tâches ordinales. En ce qui concerne les scores, 1 point est attribué pour les tâches 1, 2, 3, 4 et 2 points pour la tâche 5 dans les parties³ A et B. Cela donne un total de 6 points pour la partie A et un total de 6 points pour la partie B. Dans la partie C, 1 point est attribué pour toutes les tâches, ce qui donne un total de 6 points pour la partie C. Le score total pour le test est de 18 points. Le test n'est pas chronométré.

Tableau 2 – Types de tâches des parties A et B de l'ENUDIT 2

Types de tâches	Partie A (cardinal)	Partie B (ordinal)	Nombre d'items
Coder	A1 - Déterminer le cardinal d'une collection donnée (<i>Combien d'étoiles y a-t-il ?</i>)	B1 - Déterminer la position d'un objet (<i>À quel étage de l'immeuble se trouve le personnage ?</i>)	4 items (6, 11, 3, 4)
Décoder	A2 - Construire une collection d'un cardinal donné (<i>Peux-tu mettre 6 jetons dans un bol ?</i>)	B2 - Désigner la position d'un objet (<i>Peux-tu placer le personnage au 6e étage ?</i>)	4 items (6, 11, 3, 4)
Comparer	A3 - Comparer le cardinal de deux collections (<i>Quelle est la collection qui a le plus d'étoiles ?</i>)	B3 - Comparer deux positions (<i>Quel est le personnage qui est à l'étage le plus élevé ?</i>)	6 items (4 vs 5, 9 vs 12, 5 vs 10, 4 vs 9)
Associer	A4 - Construire une collection équipotente à une collection témoin (<i>Pose autant de jetons rouges que de jetons bleus</i>)	B4 - Reproduire une position équivalente à une position témoin (<i>Pose le personnage au même étage que dans l'immeuble témoin</i>)	4 items (6, 11, 3, 4)

² Une technique est considérée comme plus élaborée lorsque ses éléments technologiques intègrent des éléments technologiques des techniques de niveau inférieur. Par exemple, pour une tâche de comparaison, une technique de codage est considérée comme plus complexe qu'une technique de correspondance terme à terme car elle mobilise comme éléments technologiques le principe de Hume, le nombre comme mémoire d'une quantité, l'ordre des nombres ; le principe de Hume étant l'élément technologique de la correspondance terme à terme.

³ Ce choix de score a été effectué de manière arbitraire afin d'équilibrer le nombre de points par partie, en considérant que ces types de tâches sont plus complexes et plus nombreuses.

Anticiper	A5 - Résoudre des problèmes arithmétiques cardinaux/impliquant des quantités dans le champ additif (<i>Kylian a 5 crayons. Un ami lui donne 3 crayons. Combien de crayons a Kylian ?</i>)	B5 - Résoudre des problèmes arithmétiques impliquant des positions dans le champ additif (<i>Kylian est sur le 5e barreau d'une échelle, il monte de 3 barreaux. Sur quel barreau se trouve-t-il ?</i>)	6 items ($5 + 3 = ?$, $9 - 2 = ?$, $7 - 3 = ?$)
-----------	---	---	--

Tableau 3 – Types de tâches de la partie C de l'ENUDIT 2

Types de tâches	Partie C (décontextualisé)	Nombre d'items
Réciter	C0 - Réciter la comptine numérique (<i>Peux-tu compter en partant de 5 ?</i>)	3 items (compter, décompter, compter à partir de)
Associer	C3 - Nommer des codes arabes (<i>Comment s'appelle ce nombre ?</i>)	6 items (18, 25, 37, 4, 7, 13)
Ordonner	C4 - Ordonner des codes arabes (<i>Quel est le plus grand nombre ?</i>)	5 items (7, 8, 9, 10, 11 ; 12 <i>vs</i> 19, 31 <i>vs</i> 13, 3 <i>vs</i> 4, 8 <i>vs</i> 6)
Calculer	C5a - Additionner (<i>Combien font 5 plus 4 ?</i>) C5b - Soustraire (<i>Combien font 8 moins 5 ?</i>) C5c - Décomposer (<i>Donne deux cartes pour faire 6</i>)	4 items ($5 + 4$, $4 + 3$, $3 + 2$, $1 + 3$) 4 items ($8 - 5$, $9 - 3$, $5 - 3$, $4 - 1$) 4 items (8, 6, 5, 4)

IV. VALIDATION PSYCHOMÉTRIQUE DU TEST

Dans les recherches en psychologie cognitive, les outils d'évaluation de connaissances vérifient généralement plusieurs propriétés psychométriques afin de s'assurer notamment que, si l'outil est valide (c'est-à-dire si le test mesure vraiment ce qu'il prétend mesurer) et s'il est fiable (c'est-à-dire si les résultats qu'ils donnent sont stables dans le temps à divers points de mesure). Pour ce qui est de la validité, nous avons choisi d'effectuer une mesure de la validité concomitante, c'est-à-dire une comparaison entre le test et un critère de référence externe (en général un autre test déjà validé) au même moment. Pour ce qui est de la fiabilité, nous avons vérifié que l'outil est en mesure de fournir des résultats comparables entre deux passations à des temps différents (ceci assurant que les résultats obtenus ne sont pas l'effet du hasard).

1. Méthode

Les participants. La validation psychométrique du test s'est effectuée sur deux échantillons. Le premier échantillon est composé de 127 élèves de maternelle recrutés dans des écoles de la région lyonnaise (France) : 100 enfants sont issus d'écoles membres d'un Réseau d'Éducation Prioritaire (REP) (57 filles et 43 garçons avec un âge moyen de 5,6 ans et 0,2 d'écart-type) et 27 enfants sont issus d'une école favorisée (16 filles et 11 garçons avec un âge moyen de 5,8 ans et 0,2 d'écart-type). Cet échantillon a été utilisé pour mesurer la validité concomitante du test. Le second échantillon est un sous-ensemble du premier. Les enfants des écoles REP ont été recrutés à nouveau pour l'évaluation re-test et seuls 91 élèves ont pu participer. Cet échantillon a été utilisé pour mesurer la fiabilité du test.

Mesures. Pour évaluer la validité concomitante du test ENUDIT 2, nous l'avons comparé au sous-test Applied Problems de la batterie de tests Woodcock-Johnson III (WJ-III) (2001). Ce sous-test, qui a été traduit en français, évalue des connaissances mathématiques de l'école primaire et secondaire. Les premiers items font notamment appel à des concepts de base (par exemple, dénombrer, effectuer des additions et des soustractions simples). Le test n'est pas chronométré et se compose de 63 items de difficulté croissante. Après six erreurs consécutives, le test est interrompu. Ce test a été choisi parce qu'il est considéré comme un test établi et valide des connaissances globales en mathématiques, comme le montre sa large utilisation dans les études sur les connaissances mathématiques précoces (Blair et al., 2016 ; Espy et al., 2004 ; Gormley Jr et al., 2008 ; Lillard et al., 2017).

Procédure générale. Les participants ont été évalués avec le sous-test Applied Problems du WJ-III et avec le test ENUDIT 2. L'ordre dans lequel le WJ-III et l'ENUDIT 2 ont été administrés a été randomisé d'un enfant à l'autre. Les enfants ont été testés individuellement dans une pièce calme et isolée de leur école. Aucun feedback n'a été donné aux participants, que ce soit pendant ou après les tests. Chaque session de test a duré en moyenne 25 minutes (comprenant les deux tests et des pauses).

Outils d'analyse statistique. La validité concomitante de l'ENUDIT 2 a été mesurée sur l'échantillon 1, en comparant le score global de l'ENUDIT 2 avec le score obtenu au sous-test du WJ-III, à l'aide de corrélations de Pearson indépendantes. La fiabilité de l'ENUDIT 2 a été mesurée au sein de l'échantillon 2. Nous avons évalué la force de la relation entre les scores totaux obtenus lors des deux administrations du test à deux périodes différentes, à l'aide de corrélations de Pearson.

2. *Analyses descriptives et qualitatives*

Les données de 5 participants du premier échantillon et de 12 participants du second échantillon ont été retirées des analyses en raison de l'absence de résultats complets. Le premier échantillon comporte donc 122 participants et le second échantillon 79 participants.

Le tableau 3 présente les moyennes et les écarts types, pour chaque tâche du test ENUDIT 2. Le score de chaque tâche a été ramené à 1.

Tableau 4 – Statistiques descriptives pour toutes les tâches du test ENUDIT 2

	Partie A - Cardinal					Partie B - Ordinal					Partie C - Décontextualisé					
Tâches	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C0	C3	C4	C5a	C5b	C5c
Moyenne	0.85	0.81	0.85	0.55	0.32	0.77	0.77	0.86	0.49	0.22	0.49	0.58	0.74	0.51	0.17	0.26
Écart-type	0.2	0.27	0.2	0.39	0.32	0.35	0.28	0.21	0.41	0.29	0.37	0.29	0.29	0.42	0.3	0.38

Ces résultats peuvent être analysés sous plusieurs dimensions. Tout d'abord, nous remarquons des patterns de réponses similaires pour les tâches de la partie A et de la partie B : les tâches relevant de coder, décoder et comparer (A1-A2-A3, B1-B2-B3) sont réussies à plus de 75 % ; les tâches d'association (A4 et B4) sont moins bien réussies (environ 50 %) et les tâches d'anticipation (A5 et B5) sont très peu réussies (32 % et 22 % respectivement). Le moindre taux de réussite pour les tâches d'association et d'anticipation peut s'expliquer par leur nature : ce sont des tâches complexes qui nécessitent la mobilisation et l'articulation de plusieurs techniques et technologies. Nous pouvons également constater que les tâches de la partie C sont globalement moins bien réussies que celles des parties A et B, en particulier les tâches de calcul (C5). Ceci est cohérent avec l'enseignement reçu par les élèves, qui se concentre sur les usages cardinal et ordinal du nombre à l'école maternelle. Enfin, nous pouvons remarquer que les tâches d'anticipation d'une quantité (A5) et d'une position (B5) sont moins bien réussies que celles d'addition (C5a). Étant donné que les problèmes sont proposés à l'oral

et sans matériel manipulable, nous faisons l'hypothèse que la compréhension du problème, sa représentation et la recherche d'une démarche de résolution peuvent expliquer cette différence de réussite.

3. *Analyse de la validité concomitante*

La validité concomitante du test ENUDIT 2 a été évaluée en mesurant le degré de corrélation entre les scores totaux de deux tests : ENUDIT 2 et le sous-test Applied Problems du WJ-III. La corrélation de Pearson est significative et forte ($r = 0,73$, $p < 0,001$) (voir figure 1a). Cette mesure indique que les élèves les plus performants à l'ENUDIT 2 sont aussi ceux qui sont les plus performants au sous-test du WJ-III. Ainsi, le test a une excellente validité concomitante et son score composite peut être considéré comme une mesure globale des connaissances d'un élève sur le nombre au moins aussi bonne que le sous-test du WJ-III, qui est largement utilisé dans la littérature internationale.

4. *Analyse de la fiabilité*

La fiabilité du test ENUDIT 2 a été mesurée sur l'échantillon 2, c'est-à-dire les participants pour lesquels nous disposons de deux passations (l'une en février 2023 et l'autre en juin 2023). Nous avons comparé les deux scores totaux pour chaque participant ($N=79$). Le coefficient de corrélation de Pearson est significatif et fort ($r = 0,686$, $p < 0,001$) (voir figure 1b). Il indique ainsi que les élèves les plus performants à l'ENUDIT 2 à la première passation l'étaient également à la seconde passation. Ainsi le test a une bonne fiabilité test-retest.

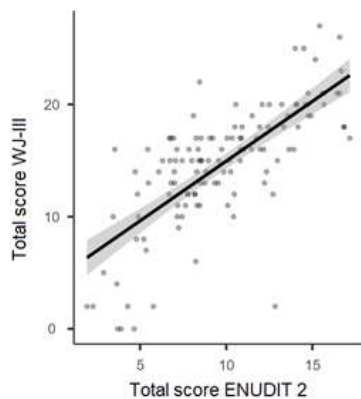


Figure 1a – Corrélation entre les scores totaux de l'ENUDIT 2 et des sous-tests du WJ-III ($N=122$)

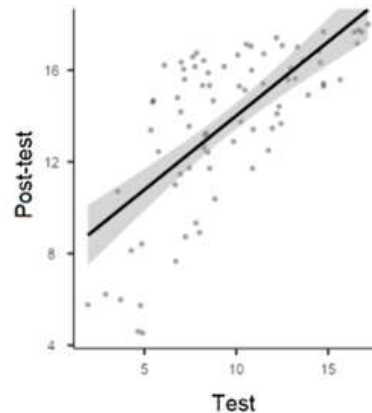


Figure 1b – Corrélation entre les deux scores du test et re-test de l'ENUDIT 2 ($N=79$)

V. CONCLUSION

Nous retenons plusieurs résultats de notre travail conjoint entre chercheurs en statistique, didactique et cognition. Le premier est l'apport des outils didactiques pour concevoir des tâches permettant d'identifier les techniques mobilisées par l'élève. La portée d'une technique permet ainsi de s'assurer, grâce au choix des valeurs de certaines variables, de l'utilisation par l'élève d'une et une seule technique pour réussir la tâche proposée. Ce travail est essentiel dans une optique d'évaluer précisément les connaissances mobilisées par les élèves. Le second résultat de notre collaboration est le recensement et la structuration des différentes techniques pour chaque type de tâches, en se référant à des travaux tant en cognition numérique qu'en didactique des mathématiques. Le troisième résultat est la validation psychométrique de l'ENUDIT 2 : il a une excellente validité concomitante et une

excellente fiabilité dans le temps. Nous obtenons ainsi un test qui répond au besoin des deux communautés : un test avec des données psychométriques, permettant d'effectuer des analyses statistiques et des analyses qualitatives fines des techniques et technologies mobilisées par les élèves. Par rapport au WJ-III qui se concentre sur le dénombrement, le transcodage et la résolution de problèmes sur des quantités, l'ENUDIT 2 permet ainsi de décrire de manière exhaustive et détaillée les connaissances des élèves sur la construction du nombre dans ses trois usages (cardinal, ordinal et décontextualisé) et pour chaque type de tâches (coder, décoder, associer, comparer, anticiper). Nous poursuivons actuellement nos réflexions, d'une part en affinant le modèle de connaissances du nombre en précisant davantage les technologies associées aux techniques, et d'autre part en cherchant à valider ce modèle didactique par des analyses confirmatoires et multidimensionnelles.

RÉFÉRENCES

- Blair, C., McKinnon, R. D. et Family Life Project Investigators. (2016). Moderating effects of executive functions and the teacher–child relationship on the development of mathematics ability in kindergarten. *Learning and Instruction*, 41, 85-93. <https://doi.org/10.1016/j.learninstruc.2015.10.001>
- Bosch, M. et Chevallard, Y. (1999). La sensibilité de l'activité mathématique aux ostensifs. *Recherches en Didactique des Mathématiques*, 19(1), 77-124.
- Briand, J. (1999). Contribution à la réorganisation des savoirs pré-numériques et numériques. *Recherches en Didactique des Mathématiques*, 19(1), 41-76.
- Cariat, A. et Margolinas, C. (2015). Le nombre comme mémoire de la position. Dans A. Bessot, C. Margolinas et L. V. Tien (dir.), *Des mathématiques à l'école maternelle* (p. 63-70). École Normale Supérieure Centrale d'Ho Chi Minh-ville. <https://hal.science/hal-01221895>
- Chaachoua, H. et Bessot, A. (2016). *Introduction de la notion de variable dans le modèle praxéologique* [Communication]. 5^e congrès pour la Théorie Anthropologique du Didactique, Castro-Urdiales, Espagne.
- Chevallard, Y. (1989, 26 juin). *Le concept de rapport au savoir: rapport personnel, rapport institutionnel, rapport officiel* [Communication]. Séminaire de didactique des mathématiques et de l'informatique, Université Joseph Fourier, Grenoble 1..
- Chevallard, Y. (1998). Analyse des pratiques enseignantes et didactique des mathématiques : L'approche anthropologique. Dans R. Noirfalise (dir.), *Actes de l'université d'été de la Rochelle, 4-11 juillet 1998* (p. 91-118). IREM de Clermont-Ferrand.
- Clements, D. H. (1999). Subitizing : What is it ? Why teach it? *Teaching Children Mathematics*, 5(7), 400-405.
- Croset, M.-C. et Gardes, M.-L. (2019). Une comparaison praxéologique pour interroger l'enseignement du nombre dans l'institution Montessori. *Recherches en Didactique des Mathématiques*, 39(1), 51-96.
- Croset, M.-C. et Gardes, M.-L. (2020). Une carte des connaissances pour la construction du nombre en maternelle. *Revue de Mathématiques pour l'École*, 233, 117-127. <https://doi.org/10.26034/vd.rm.2020.1742>
- Dehaene, S. (2008). Psychologie cognitive expérimentale. *L'annuaire du Collège de France. Cours et travaux*, 108, 277-301.

- Dehaene, S. (2018). *Apprendre ! Les talents du cerveau, le défi des machines*. Odile Jacob.
- Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A. et Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology*, 26(1), 465-486.
- Gelman, R., Gallistel, C. R. et Gelman, R. (2009). *The child's understanding of number*. Harvard University Press.
- Gormley, W. T. Jr, Phillips, D. et Gayer, T. (2008). Preschool programs can boost school readiness. *Science*, 320(5884), 1723-1724.
- Halberda, J. et 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. <https://doi.org/10.1037/a0012682>
- Lillard, A., Heise, M., Richey, E., Tong, X., Hart, A. et Bray, P. (2017). Montessori preschool elevates and equalizes child outcomes: A longitudinal study. *Frontiers in Psychology*, 8, Article 1783. <https://doi.org/10.3389/fpsyg.2017.01783>
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S. et Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, 116(1), 33-41. <https://doi.org/10.1016/j.cognition.2010.03.012>
- Woodcock, R. W., McGrew, K. S. et Mather, N. (2001). *Woodcock-Johnson tests of achievement*. Riverside Publishing.
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358(6389), 749-750. <https://doi.org/10.1038/358749a0>

ANNEXE 1.

UN MODÈLE DE CONNAISSANCES EN JEU LORS DE LA CONSTRUCTION DU CONCEPT DE NOMBRE

