

**Dépistage des difficultés mathématiques : validation et normalisation franco-qubécoise du Tempo Test Rekenen et du Numeracy Screener version française**

Anne LAFAY \*, Helena Patricia OSANA \*\*, Sarah MICHAUD \*\*\*, Nadia NOSWORTHY \*\*\*\*

\* Ph-D, orthophoniste, membre de l'Ordre des Orthophonistes et Audiologistes du Québec, chercheuse postdoctorale Banting, Mathematics Teaching and Learning Lab, Concordia University, Montréal, Canada

\*\* Ph-D, professeure, Mathematics Teaching and Learning Lab, Concordia University, Montréal, Canada

\*\*\* orthophoniste, Nice, France

\*\*\*\* Ph-D, professeure, Department of Graduate Psychology and Counseling, Andrews University, USA

**Auteure de correspondance :**  
lafay\_anne@yahoo.fr

**Résumé :**

L'objectif est d'évaluer les qualités psychométriques de deux tests mathématiques dans le but de développer des normes franco-québécoises pour le dépistage et l'identification des enfants à risque de présenter un trouble des apprentissages en mathématiques : le Tempo Test Rekenen (TTR) visant l'évaluation des capacités de fluence arithmétique et le Numeracy Screener version française (NS-f) visant l'évaluation des habiletés de base du traitement du nombre et des numérosités. Un total de 695 enfants franco-québécois scolarisés de la maternelle à la 6<sup>ème</sup> année de primaire a été évalué. Les résultats ont montré une bonne cohérence interne, une bonne validité prédictive et une bonne validité de construit. Ils suggèrent ainsi que le présent dispositif de dépistage comportant le NS-f et le TTR, a une bonne validité pour le dépistage et l'identification des enfants à risque de présenter un trouble des apprentissages en mathématiques.

**Mots clés :** Dépistage, Mathématiques, Dyscalculie, Validité, Normes

**Identification of mathematical difficulties: validation and French-Quebec normalization of the Tempo Test Rekenen and the Numeracy Screener French version**

**Summary:**

The aim is to investigate the psychometric properties of two mathematical tools in order to receive French-Quebec normative data for the screening and identification of children at-risk for mathematical learning disabilities: the Tempo Test Rekenen (TTR) that assesses arithmetic fluency and the French version of the Numeracy Screener (NS-f) that assesses the basic skills of number and numerosities processing. These tests were administered to a total of 695 francophone children in Quebec from kindergarten to 6<sup>th</sup> grade. Results showed good internal consistency, predictive validity, and construct validity. Results thus suggest that the set of NS-f and TTR has good validity for screening and identifying children at-risk for mathematical learning disabilities.

**Keywords:** Screening, Mathematics, Dyscalculia, Validity, Normative data

## ----- INTRODUCTION -----

Jusqu'à 10% des enfants d'âge scolaire présentent des difficultés en mathématiques (Dirks, Spyer, van Lieshout & de Sonnevile, 2008). Or, les capacités mathématiques sont fortement prédictives de la réussite scolaire du jeune apprenant (Duncan et al., 2007). Elles sont aussi fortement associées au statut socio-économique des citoyens (Ritchie & Bates, 2013) et à la qualité de vie (Bynner & Bates Parsons, 1997). Les activités mathématiques représentent des défis importants pour ces enfants. La présente étude a pour objectif d'évaluer les qualités psychométriques de deux tests mathématiques dans le but ultime de développer des normes franco-québécoises qui seront utiles au dépistage et à l'identification des enfants à risque de présenter un trouble des apprentissages en mathématiques (DSM-5, 2013 ; version française, 2016).

Le modèle de la Réponse À l'Intervention (RàI ; National Center on Response to Intervention, NCRTI, 2010) est un cadre d'organisation de services du système scolaire se mettant peu à peu en place au Québec (Ministère de l'Éducation, du Loisir et du Sport, MELS, 2011 ; Ministère de l'Éducation de l'Enseignement supérieur et de la Recherche, MEES, 2015). Il comprend des procédés d'évaluation et d'intervention s'appuyant sur les données actuelles de la recherche dans un système axé sur la prévention. Il est ainsi constitué de trois paliers destinés à maximiser la réussite des jeunes. Le palier 1, niveau universel, comprend un enseignement de qualité dans lequel sont utilisés des programmes et des approches pédagogiques en salle de classe auprès de l'ensemble des élèves. Le palier 2, enseignement supplémentaire ciblé, est constitué d'interventions, le plus souvent en petits groupes, davantage ajustées en fonction des forces et besoins des élèves pour qui l'enseignement en salle de classe ne s'est pas avéré suffisant pour assurer un rythme adéquat des apprentissages. Le palier 3, niveau d'intervention intensif, comprend des interventions intensives, adaptées aux besoins spécifiques des élèves, généralement offertes individuellement. Le présent article cible tout particulièrement le palier 2 pour lequel des évaluations de dépistage sont nécessaires pour identifier les enfants à risque de présenter un trouble des apprentissages en mathématiques et nécessitant une aide supplémentaire.

Le dépistage est un défi majeur pour les enseignants et les professionnels pour parvenir à identifier les enfants à risque de présenter un trouble des apprentissages en mathématiques. Avoir à disposition des outils valides, fidèles et normés pour la population cible est indispensable pour permettre une identification la plus objective possible. Il existe quelques outils d'évaluation mathématique normalisés pour la population franco-québécoise (Lafay, St-Pierre & Macoir, 2014 ; Lafay & Cattini, 2018). Par exemple, le WIAT-II (Wechsler, 2005) a été traduit de l'anglais vers le français mais ce test nécessite un temps long de passation individuelle. Le manque d'outils contribue fortement à la situation de « sous-identification » des enfants présentant un trouble des apprentissages en mathématiques. Le présent article cible ainsi deux outils de dépistage des difficultés mathématiques : le Tempo Test Rekenen (TTR ; De Vos, 1992) et le Numeracy Screener version française (NS, Nosworthy, 2013 ; NS-f, Lafay, Archambault, Vigneron & Nosworthy, 2018). Ceux-ci ont l'intérêt d'être de passation très courte et collective. De plus, les temps de correction et de cotation sont également très courts et très faciles pour le professionnel. Enfin, ces deux outils ciblent des compétences particulièrement primordiales du développement mathématique : respectivement les habiletés de fluence arithmétique et les habiletés de base du traitement du nombre et des numérosités.

Le TTR (De Vos, 1992) est un test de fluence arithmétique néerlandais et permet une évaluation graduée des quatre opérations arithmétiques. Il a été sélectionné pour le présent dispositif de dépistage pour plusieurs raisons.

Premièrement, le calcul est une capacité indispensable pour bien évoluer dans le cheminement scolaire et pour réaliser diverses tâches de la vie quotidienne. D'ailleurs, la résolution d'additions est prédictive des habiletés mathématiques générales chez des enfants de 1<sup>ère</sup> à 5<sup>ème</sup> année de primaire (Geary, 2011). Deuxièmement, les difficultés de calcul sont un des symptômes principaux identifiés dans le trouble des apprentissages en mathématiques (DSM-5, 2013 ; version française, 2016). Troisièmement, le TTR a déjà montré qu'il possédait des propriétés psychométriques importantes. En effet, une première étude (Lafay, St-Pierre & Macoir, 2015) auprès de 77 enfants franco-québécois âgés de 8 ans à 9 ans et scolarisés en 3<sup>ème</sup> année de primaire dans la région de Québec dans le cadre d'un enseignement régulier a montré que la performance au TTR est corrélée avec le jugement du niveau scolaire et du niveau mathématique évalué par l'enseignant et avec celui évalué par le parent d'une part, ainsi qu'avec un test d'évaluation des capacités mathématiques globales d'autre part (à savoir, le Zareki-R ; Dellatolas & von Aster, 2006). Ces résultats suggèrent donc de bonnes validités prédictive et concomitante pour le TTR et justifient de poursuivre les études.

Le NS-f (Lafay, Archambault et al., 2018) est la version française du test Numeracy Screener (Nosworthy, 2013). Le NS-f a été sélectionné pour le présent dispositif de dépistage pour plusieurs raisons. Premièrement, il s'agit d'un outil spécialement conçu pour le dépistage des difficultés mathématiques, en ciblant tout particulièrement les habiletés de base du traitement du nombre et des numérosités. Ce test est en effet une mesure de la capacité de l'enfant à comprendre les nombres et les numérosités (Nosworthy, 2013). Ce test s'accorde parfaitement avec les modèles théoriques du développement du nombre et du calcul. Quatre moments importants dans le développement du traitement du nombre sont décrits par von Aster et Shalev (2007). Le développement mathématique s'appuie initialement sur ce qu'on appelle le *sens du nombre* (Dehaene, 2010) ou le *module nombre* (Butterworth, 1999). La représentation analogique (c.-à-d. non-symbolique) des nombres, à savoir les représentations numériques mentales, est fondamentale pour le développement des compétences numériques. L'enfant acquiert par la suite le code numérique oral et apprend le code numérique arabe pour parvenir à des représentations numériques dites matures modélisées par une ligne numérique. Deuxièmement, les difficultés de sens et traitement du nombre sont un des symptômes principaux identifiés dans le trouble des apprentissages en mathématiques (DSM-5, 2013 ; version française, 2016). D'ailleurs, le sens du nombre est au cœur des hypothèses explicatives de la dyscalculie (Butterworth, 2005 ; von Aster & Shalev, 2007 ; Wilson & Dehaene, 2007 ; Lafay, St-Pierre, & Macoir, 2019), de même que le déficit d'accès au sens du nombre (Lafay, Macoir, & St-Pierre, 2018 ; Lafay, St-Pierre, & Macoir, 2017 ; Noël & Rousselle, 2011). Troisièmement, des résultats expérimentaux attestent déjà de la validité du Numeracy Screener. Des corrélations ont été identifiées entre la performance au Numeracy Screener et les capacités arithmétiques dans la population des enfants canadiens anglophones (Nosworthy, Zheng, & Ansari, 2014) suggérant une bonne validité prédictive. De même, la performance au Numeracy Screener chez 160 enfants de 1<sup>ère</sup> année de primaire à 3<sup>ème</sup> année de primaire était en mesure d'expliquer les différences interindividuelles en termes de capacité arithmétique (Nosworthy, Bugden, Archibald, Evans, & Ansari, 2013). De plus, une analyse descriptive des performances obtenues au NS-f par 19 enfants au profil varié (Lafay et al., 2018) a montré que les enfants ne présentant pas de plainte en mathématiques ni même de trouble développemental du langage obtenaient effectivement des scores au NS-f au-dessus du percentile 35 confirmant ainsi l'absence de difficultés de traitement du nombre et des quantités. Inversement, les enfants suivis

en mathématiques ou les enfants suivis en langage pour lesquels un trouble des apprentissages en mathématiques était suspecté par les orthophonistes<sup>1</sup> obtenaient effectivement des scores au NS-f sous le percentile 10 suggérant un trouble des apprentissages en mathématiques ou sous le percentile 35 suggérant, à tout le moins, des difficultés mathématiques (Green & Gallagher, 2014), confirmant ainsi la présence de difficultés de traitement du nombre et des quantités. Ces résultats suggèrent ainsi une bonne sensibilité et une bonne spécificité du test.

## ----- OBJECTIFS -----

L'objectif de la présente étude est d'évaluer les qualités psychométriques de deux tests mathématiques, le TTR et le NS-f, dans le but ultime de développer des normes franco-québécoises qui seront utiles au dépistage et à l'identification des enfants à risque de présenter un trouble des apprentissages en mathématiques. Pour cela, un groupe d'enfants franco-québécois scolarisés de la maternelle à la 6<sup>ème</sup> année de primaire a été évalué en classe au moyen des tests TTR et NS-f. Leurs performances ont tout d'abord été comparées à celles obtenues par des enfants franco-québécois testés en 2014, par des enfants canadiens anglophones testés en 2013 et par des enfants français testés en 2018 dans l'objectif de démontrer l'importance de développer des normes spécifiques au contexte franco-québécois. Ensuite, les performances obtenues par les enfants franco-québécois de la présente étude ont été analysées dans l'objectif d'examiner la validité du TTR et du NS-f. Des analyses en fonction du niveau de classe, du genre, de l'école et des conditions de passation ont également été effectuées pour décider si les normes à produire devaient prendre en considération ces critères. Enfin, les normes sont présentées.

## ----- MÉTHODE -----

### 1. Participants

Deux écoles franco-québécoises ont participé à la présente étude. L'école E1 de la commission scolaire des Hauts Cantons et l'école E2 de la commission scolaire de La Capitale. Selon les indices de défavorisation du MEES (2016-2017)<sup>2</sup>, l'école E1 a un indice de défavorisation la situant au décile 7 et l'école E2 a un indice de défavorisation la situant au décile 8. Les élèves de 2<sup>ème</sup> et 3<sup>ème</sup> cycle de l'école E1 ont participé au projet *CalculoManiac* visant à développer les stratégies de calcul mental tout au long de l'année en cours.

Au total, 695 enfants scolarisés de la maternelle à la 6<sup>ème</sup> année de primaire ont été évalués. Toutefois, deux élèves ont été retirés des analyses car leur score était de 0. Ils avaient été identifiés comme ayant des difficultés de compréhension par les enseignants et les orthopédagogues, et nous pouvons donc supposer qu'ils n'ont tout simplement pas compris la consigne. L'échantillon final se compose donc de 693 enfants : 391 élèves de l'école E1 (l'école E1 n'a pas identifié le genre des élèves) et 302 élèves de l'école E2, 139 filles et 163 garçons. Une analyse khi<sup>2</sup> a été réalisée pour analyser la répartition des genres par niveau de classe et montre que la répartition en genre est parfaitement équivalente selon chaque niveau de classe

<sup>1</sup> L'étude de Michaud (2018) a été réalisée en France, pays dans lequel l'évaluation et l'intervention auprès d'enfants présentant un trouble des apprentissages en mathématiques font partie des compétences des orthophonistes.

<sup>2</sup> Les écoles sont classées sur une échelle allant de 1 à 10, le rang 1 étant considéré comme le plus favorisé et le rang 10 comme le plus défavorisé.

( $\chi^2(6, N = 307) = 3.627, \varphi = .109, p = .727$ ). Une analyse khi<sup>2</sup> a aussi été réalisée pour analyser la répartition des niveaux de classe par école et montre que la répartition n'est pas parfaitement équivalente ( $\chi^2(6, N = 693) = 19.081, \varphi = .166, p = .004$ ). Les deux écoles comptent un pourcentage équivalent d'élèves de 1<sup>ère</sup> et 4<sup>ème</sup> années. Cependant, l'école E1 compte un pourcentage plus grand d'élèves de 5<sup>ème</sup> et 6<sup>ème</sup> années, alors que l'école E2 compte un pourcentage plus grand d'élèves de maternelle, 2<sup>ème</sup> et 3<sup>ème</sup> années.

Une analyse de variance (ANOVA) a été réalisée pour évaluer l'effet de trois variables indépendantes (école, genre et niveau de classe) sur la variable dépendante de l'âge dans l'objectif d'éliminer la variable confondante de l'âge dans les analyses futures. Les résultats ne montrent aucun effet principal de l'école ( $F(1,288) = 0.084, p = .772$ ) ni de genre ( $F(1,288) = 0.278, p = .598$ ), mais un effet principal du niveau de classe sur l'âge ( $F(6,288) = 929.671, p < .001, \eta_p^2 = .951$ ). Aucune interaction n'est observée. De manière attendue, les élèves des deux écoles ont le même âge, les garçons et les filles ont le même âge mais l'âge augmente d'une année sur l'autre. Le tableau 1 résume l'effectif et l'âge par niveau scolaire.

**Tableau 1**

*Caractéristiques démographiques par école et par niveau de classe*

	Effectif	Âge (moyenne, en mois)	Âge (écart-type, en mois)
Maternelle	97	73,16	3,65
1 <sup>e</sup> année	101	85,81	4,27
2 <sup>e</sup> année	118	97,99	6,33
3 <sup>e</sup> année	96	110,24	5,73
4 <sup>e</sup> année	97	122,71	5,41
5 <sup>e</sup> année	91	134,10	5,16
6 <sup>e</sup> année	93	145,42	4,75

## 2. Mesures

### a. Tempo Test Rekenen (TTR)

Le TTR est un test de fluence arithmétique. L'enfant dispose d'une feuille qui contient cinq colonnes de calculs écrits. Le test est administré selon les consignes suivantes : « Réalise le plus d'additions / soustractions / multiplications / divisions / opérations mélangées possible en une minute » respectivement pour chaque colonne. Chaque colonne comporte 40 calculs écrits. Les calculs sont ordonnés du plus simple (ex.  $1+1=$ \_\_ ;  $2 \times 2=$ \_\_), exigeant un recours aux faits arithmétiques en mémoire (Campbell & Oliphant, 1992) ou des procédures automatiques et très rapides de calcul (Thevenot, 2017), au plus complexe (ex.  $54+27=$ \_\_ ;  $5 \times 17=$ \_\_), exigeant le recours à des procédures de calcul mental. Un score pour chaque type d'opération est calculé en attribuant un point à chaque calcul réussi (score minimal de 0 ; score maximal de 40). Un score " Petit total " est calculé en additionnant le sous-score correspondant aux additions et soustractions (score minimal de 0 ; score maximal de 80). Un score " Grand total " est calculé en additionnant chaque sous-score correspondant aux cinq colonnes (score minimal de 0 ; score maximal de 200).

### b. Numeracy Screener version française (NS-f)

Le NS-f (Lafay, Archambault et al., 2018) est la version française du test Numeracy Screener (Nosworthy, 2013). Il s'agit d'un outil de dépistage des difficultés mathématiques, en particulier des habiletés de base du traitement du nombre et des numérosités. Le NS-f est

composé de deux parties : une section Symbolique visant l'évaluation du traitement des nombres arabes (4 pages, 56 items) et une partie Non-Symbolique visant l'évaluation du traitement des quantités analogiques (4 pages, 56 items). L'enfant dispose d'une feuille sur laquelle il voit des paires de nombres arabes (partie Symbolique) ou d'ensembles de points (partie Non-Symbolique). La partie Symbolique est administrée selon la consigne « Je veux que tu regardes ces nombres et que tu décides quel nombre est le plus grand ». La partie Non-symbolique est administrée selon la consigne « Je veux que tu regardes ces points et que tu décides dans quelle case il y a le plus de points ». La durée d'administration du test est de 2 à 4 minutes (selon l'âge). Pour chacun des subtests, l'élève dispose d'un temps imparti pour tenter de réaliser un maximum d'items. Si tous les items sont effectués avant la fin du temps imparti, le temps mis par l'enfant est annoté sur l'épreuve. La forme A du test propose la partie symbolique puis la partie non symbolique alors que la forme B propose la partie non symbolique puis la partie symbolique. La variable dépendante est le ratio du nombre d'items corrects obtenu en une minute selon la formule suivante : [nombre d'items corrects / temps total \* 60 secondes].

### 3. Procédure générale

La collecte des données a été réalisée par un binôme constitué d'orthopédagogues ou d'une conseillère pédagogique et d'une orthopédagogue. Les tests ont été administrés collectivement dans la classe respective de chaque enfant. La passation complète (présence en classe de la présentation de la recherche à la récupération des documents) a duré environ 20 minutes selon les classes. Bien que le temps de passation de chaque subtest soit très court (2 à 4 minutes pour le NS-f et 2 à 5 minutes pour le TTR), l'organisation d'une passation collective a pris plus de temps qu'initialement prévu.

Les enfants de maternelle n'ont passé que le NS-f. Les enfants de 1<sup>ère</sup> et 2<sup>ème</sup> année ont passé le NS-f ainsi que les colonnes d'additions et de soustractions du TTR. Les enfants de 3<sup>ème</sup> année ont passé le NS-f et le TTR dans son intégralité. Les enfants de 4<sup>ème</sup> à 6<sup>ème</sup> année ont passé le NS-f et le TTR au complet dans l'école E2 alors qu'ils ont passé uniquement le TTR dans l'école E1.

L'ordre de passation a été pseudo-randomisé. La passation étant collective dans chaque classe, il n'a pas été possible de randomiser l'ordre de passation des tests pour chaque enfant. En revanche, un ordre différent entre le TTR et le NS-f ainsi qu'entre la forme A et B du NS-f a été déterminé, au hasard, pour chaque classe. Concernant les maternelles, 65 élèves ont passé la forme A du NS-f et 32 élèves ont passé la forme B du NS-f. Concernant les classes de 1<sup>ère</sup> à 6<sup>ème</sup> année ayant passé les deux tests, 158 élèves ont passé le TTR puis la forme A du NS-f, 173 élèves ont passé le TTR puis la forme B du NS-f, 124 élèves ont passé la forme A du NS-f puis TTR et 38 élèves ont passé la forme B du NS-f puis TTR.

### 4. Procédure d'analyse

Deux stagiaires étudiantes en orthophonie (Institut des Sciences et Techniques de Réadaptation, université Claude Bernard Lyon 1 ; en stage d'initiation à la recherche à l'université Concordia) ont effectué la correction de chaque copie d'élèves. Ensuite, ces mêmes étudiantes ainsi qu'une assistante de recherche (université Concordia) ont réalisé la saisie de données sur format informatique. Chaque copie a été corrigée et chaque résultat saisi par deux personnes différentes, de sorte que la personne entrant les données était en charge de vérifier la correction

effectuée par la première personne. Ensuite, les analyses statistiques (descriptives et inférentielles) ont été réalisées par la première auteure du projet à l'aide du logiciel SPSS version 23.

Les performances d'un groupe d'enfants franco-québécois ont tout d'abord été comparées à celles obtenues par des enfants franco-québécois, testées en 2014, par des enfants canadiens anglophones, testées en 2013, et par des enfants français, testées en 2018, au moyen d'analyses de variance (ANOVA). Ensuite, les données collectées dans le groupe d'enfants franco-québécois ont été analysées pour obtenir les indices de cohérence interne (alpha de Cronbach, corrélations, analyse de régression). Aussi, des analyses de comparaison entre les différents niveaux, les genres, les conditions de passation et les écoles ont été effectuées au moyen d'ANOVA. Enfin, les normes en moyennes et écarts-types ou percentiles ont été produites.

## ----- RÉSULTAT -----

### 1. Comparaison des performances au TTR et NS-f avec celles d'autres périodes temporelles et régions géographiques

#### a. Comparaison des scores au TTR entre deux périodes temporelles

Les scores au TTR de 77 enfants franco-québécois scolarisés en 3<sup>ème</sup> année de primaire en 2013-2014 (données issues de Lafay et al., 2015) et ceux de 96 enfants franco-québécois scolarisés en 3<sup>e</sup> année de primaire en 2017-2018 (données issues de la présente étude) ont été comparés. Une série d'analyses de variance (ANOVA) a été réalisée pour évaluer l'effet du temps de passation sur la performance au TTR. Les résultats ne montrent aucun effet principal du temps sur le score Addition ( $F(1,171) = 10.763, p = .627$ ), sur le score Soustraction ( $F(1,171) = 2.327, p = .686$ ), sur le score Multiplication ( $F(1,171) < 0.001, p = .985$ ), sur le score Opérations mêlées ( $F(1,171) = 0.062, p = .803$ ), sur le Grand total ( $F(1,171) = 1.898, p = .170$ ), suggérant que les deux groupes ne diffèrent pas. En revanche, un effet principal du temps est noté sur le score Division ( $F(1,469) = 53.208, p < .001, \eta_p^2 = .237$ ) montrant que les enfants testés en 2017-2018 obtiennent de meilleures performances que ceux testés en 2013-2014. Cette différence pourrait s'expliquer par le fait que les enfants de 2013-2014 ont été évalués entre février et juin alors que les enfants de 2017-2018 ont été évalués en mai. Ce résultat suggère une bonne fidélité temporelle mais la différence dans le score Division justifie de produire des nouvelles normes pour l'ensemble des niveaux de classe au même moment de l'année.

#### b. Comparaison des performances au NS-f entre le Québec et l'Ontario

Les scores de 160 enfants d'Ontario (données issues de Nosworthy, 2013) et 305 enfants du Québec de 1<sup>ère</sup>, 2<sup>ème</sup> et 3<sup>ème</sup> années ont été analysés. Notons que les enfants d'Ontario et du Québec ont été évalués durant une période similaire (en mai et juin pour les enfants d'Ontario, en mai pour les enfants du Québec). Tout d'abord, une série d'ANOVAs a été réalisée pour évaluer l'effet de province et du niveau de classe sur l'âge. Les résultats montrent un effet principal de province ( $F(1,469) = 70.172, p < .001, \eta_p^2 = .130$ ), un effet principal du niveau de classe ( $F(1,469) = 690.512, p < .001, \eta_p^2 = .746$ ) mais aucune interaction Province x Niveau de classe ( $F(2,469) = .237, p = .762$ ). Les enfants du Québec sont plus vieux que les enfants d'Ontario, de 4 à 5 mois selon le niveau de classe. L'âge est pris comme covariable des analyses suivantes.

Une série d'analyses de covariance (ANCOVA) a été réalisée pour évaluer l'effet de province et du niveau de classe sur les performances obtenues au NS-f (score Non-Symbolique, score Symbolique et score total) pour les classes de 1<sup>ère</sup>, 2<sup>ème</sup> et 3<sup>ème</sup> année. Notons que les franco-québécois ont passé le NS-f alors que les ontariens avaient passé la forme initiale (en anglais) du NS. Concernant le score Non-Symbolique, les résultats ne montrent aucun effet principal de province ( $F(1,468) = 0.116, p = .734$ ) mais un effet principal attendu du niveau de classe ( $F(1,468) = 8.492, p < .001, \eta_p^2 = .035$ ) et une interaction Province x Niveau de classe ( $F(2,468) = 4.971, p < .001, \eta_p^2 = .021$ ). Des analyses de comparaison (*t*-tests avec correction Bonferroni) montrent que les élèves de 1<sup>ère</sup> année d'Ontario sont meilleurs que ceux du Québec alors qu'aucune différence n'est observée pour les élèves de 2<sup>ème</sup> et 3<sup>ème</sup> années. Concernant le score Symbolique, les résultats montrent un effet principal de province ( $F(1,468) = 19.151, p < .001, \eta_p^2 = .039$ ) et un effet principal attendu du niveau de classe ( $F(1,468) = 7.787, p < .001, \eta_p^2 = .032$ ) mais aucune interaction Province x Niveau de classe ( $F(2,468) = 1.813, p = .164$ ). Les enfants d'Ontario sont moins performants que ceux du Québec. Concernant le score total, les résultats montrent un effet principal de la province ( $F(1,468) = 6.573, p = .011, \eta_p^2 = .014$ ), un effet principal attendu du niveau de classe ( $F(1,468) = 10.030, p < .001, \eta_p^2 = .041$ ) et une interaction Province x Niveau de classe ( $F(2,468) = 3.816, p = .023, \eta_p^2 = .016$ ). Des analyses de comparaison (*t*-tests avec correction Bonferroni) ne montrent aucune différence pour les élèves de 1<sup>ère</sup> année, alors que les 2<sup>ème</sup> et 3<sup>ème</sup> années du Québec sont meilleurs que ceux d'Ontario. Les résultats suggèrent que la fidélité entre provinces et entre langues à l'intérieur du Canada n'est pas toujours respectée.

### c. Comparaison des performances au NS-f entre le Québec et la France

Les scores de 215 enfants de France de CM2 et 6<sup>ème</sup> collège (correspondant aux 5<sup>ème</sup> et 6<sup>ème</sup> années ; données issues de Michaud, 2018) et 124 enfants du Québec de 5<sup>ème</sup> et 6<sup>ème</sup> années ont été analysés. Notons que les enfants de France ont été évalués entre décembre et février alors que les enfants du Québec ont été évalués en mai. Tout d'abord, une ANOVA a été réalisée pour évaluer l'effet de pays et du niveau de classe sur l'âge. Les résultats montrent un effet principal de pays ( $F(1,332) = 254.727, p < .001, \eta_p^2 = .434$ ), un effet principal du niveau de classe ( $F(1,332) = 654.559, p < .001, \eta_p^2 = .663$ ) mais aucune interaction Pays x Niveau de classe ( $F(2,332) = .088, p = .767$ ). Comme attendu, les enfants du Québec sont plus vieux que les enfants de France de 6 mois en moyenne. L'âge est donc pris comme covariable des analyses suivantes.

Une série d'ANCOVAs a été réalisée pour évaluer l'effet de pays et du niveau de classe sur les performances obtenues au NS-f (score Non-Symbolique, score Symbolique et score total) pour les classes de 5<sup>ème</sup> et 6<sup>ème</sup> années. Notons que les franco-québécois ont passé le NS-f alors que les français ont passé une forme similaire traduite (non publiée) par Michaud (2018). Concernant le score Non-Symbolique, les résultats montrent un effet principal de pays ( $F(1,331) = 21.333, p < .001, \eta_p^2 = .061$ ) et un effet principal du niveau de classe ( $F(1,331) = 9.899, p = .002, \eta_p^2 = .029$ ) mais aucune interaction Pays x Niveau de classe ( $F(2,331) = 1.683, p = .195$ ). Concernant le score Symbolique, les résultats montrent un effet principal de pays ( $F(1,331) = 135.716, p < .001, \eta_p^2 = .291$ ) et un effet principal du niveau de classe ( $F(1,331) = 7.702, p = .006, \eta_p^2 = .023$ ) mais aucune interaction Pays x Niveau de classe ( $F(2,331) = 0.860, p = .355$ ). Concernant le score total, les résultats montrent un effet principal de pays ( $F(1,331) = 99.115, p < .001, \eta_p^2 = .230$ ) et un effet principal du niveau de classe ( $F(1,331) = 11.679, p < .001, \eta_p^2 = .034$ ) mais aucune interaction Pays x Niveau de classe ( $F(2,331) = 1.607, p = .206$ ). Pour l'ensemble des trois scores (score Non-Symbolique, score Symbolique et score total), les

jeunes français obtiennent des performances inférieures à celles des jeunes franco-québécois quel que soit le niveau de classe. Les résultats suggèrent que la fidélité entre pays n'est pas toujours respectée.

## d. En résumé

Les performances obtenues par le groupe d'enfants franco-québécois diffèrent de celles obtenues par un groupe d'enfants franco-québécois en 2013-2014 (TTR), de celles obtenues par un groupe d'enfants canadiens anglophones en 2013 (NS-f) et de celles obtenues par un groupe d'enfants français en 2017-2018 (NS-f). Ces différences démontrent ainsi l'importance de développer des normes spécifiques au contexte québécois francophone.

## 2. Validité du TTR et du NS-f

### a. Relation entre les subtests du TTR

Tout d'abord, une analyse de la cohérence interne du TTR a été réalisée. Les résultats montrent que l'alpha de Cronbach est  $\alpha = .946$ . De plus, une analyse de corrélations de Pearson a été réalisée pour évaluer la cohérence interne inter-subtests du TTR visant à évaluer les habiletés de fluence arithmétique. Les résultats mettent en évidence que les cinq subtests sont fortement corrélés entre eux et aux deux scores totaux (Petit total et Grand total). Le tableau 2 indique l'ensemble des corrélations. Les résultats suggèrent une bonne cohérence interne du TTR.

**Tableau 2**

*Corrélation entre les scores des subtests du TTR*

	(2)	(3)	(4)	(5)	(6)	(7)
<b>Addition (1)</b>	0,890**	0,796**	0,709**	0,817**	0,974**	0,906**
<b>Soustraction (2)</b>		0,709**	0,684**	0,803**	0,971**	0,875**
<b>Multiplication (3)</b>			0,803**	0,851**	0,791**	0,922**
<b>Division (4)</b>				0,822**	0,731**	0,889**
<b>Mêlé (5)</b>					0,850**	0,948**
<b>Petit total (6)</b>						0,935**
<b>Grand total (7)</b>						

\*\*  $p < 0,001$

Note. Les corrélations ont été calculées pour un total oscillant entre 375 enfants (pour les corrélations impliquant les scores de multiplication, division et opérations mêlées) à 594 enfants.

### b. Relation entre les subtests du NS-f

Tout d'abord, une analyse de la cohérence interne du NS-f a été réalisée. Les résultats montrent que l'alpha de Cronbach est  $\alpha = .796$ . De plus, une analyse de corrélations de Pearson a été réalisée pour évaluer la cohérence interne inter-subtests du NS-f visant à évaluer les habiletés de base de traitement du nombre. Les résultats mettent en évidence une forte corrélation entre le score du NS-f Non-Symbolique et le score du NS-f Symbolique ( $N = 588$ ,  $r = .715$ ,  $p < .001$ ), entre le score du NS-f Non-Symbolique et le score total du NS-f ( $N = 588$ ,  $r = .892$ ,  $p < .001$ ) ainsi qu'entre le score du NS-f Symbolique et le score total du NS-f ( $N = 588$ ,  $r = .953$ ,  $p < .001$ ). Les résultats suggèrent une bonne cohérence interne du NS-f.

### c. Relation entre le TTR et le NS-f

Une analyse de corrélations de Pearson a été réalisée pour évaluer la cohérence interne du présent dispositif de dépistage (TTR et NS-f). Les résultats mettent en évidence que les deux tests sont fortement corrélés entre eux. Le tableau 3 indique l'ensemble des corrélations. Les résultats suggèrent une bonne cohérence interne de notre dispositif global de dépistage (TTR et NS-f).

**Tableau 3**

*Corrélation entre les scores du NS-f et du TTR*

	TTR Petit total	TTR Grand total
NS-f Non symbolique	0,574**	0,373**
NS-f Symbolique	0,790**	0,731**
NS-f Total	0,765**	0,669**

\*\*  $p < 0,001$

Note. Les corrélations ont été calculées pour un total de 270 ou 489 enfants ayant fait les deux tests.

De plus, une analyse de régression pas à pas a été réalisée pour évaluer si les scores du NS-f Non-symbolique et Symbolique sont des prédicteurs des scores du TTR Petit total et Grand total. Les prédicteurs intégrés au modèle sont l'âge, le score NS-f Non-Symbolique et le score NS-f Symbolique. La première analyse concernant le TTR Petit total montre que l'âge explique 63.9 % de la variance (voir tableau 4). De plus, le score au NS-f Non-Symbolique contribue significativement à l'amélioration de l'explication de la variabilité du TTR Petit total de sorte que le modèle « âge et NS-f Non-Symbolique » explique 66.5 % de la variance. Enfin, le score au NS-f Symbolique contribue significativement à l'amélioration de l'explication de la variabilité du TTR Petit total de sorte que le modèle « âge et NS-f Non-Symbolique et NS-f Symbolique » explique 71.9 % de la variance, avec l'âge et NS-f Symbolique qui sont des prédicteurs significatifs.

**Tableau 4**

*Résumé de l'analyse de régression hiérarchique pour le TTR Petit total*

Variable	R	R <sup>2</sup>	R <sup>2</sup> ajusté	$\beta$	$t$
Étape 1					
Âge	.800	.639	.639	.800	29.299**
Étape 2	.815	.665	.663		
Âge				.695	22.077**
NS-f Non-Symbolique				.190	6.033**
Étape 2	.848	.719	.717		
Âge				.473	12.831**
NS-f Non-Symbolique				.041	1.266
NS-f Symbolique				.403	9.675**

\* $p < .05$ , \*\* $p < .001$

La deuxième analyse concernant le TTR Grand total montre que l'âge explique 70.2 % de la variance ( $p < .001$ ) (voir tableau 5). De plus, le score au NS-f Non-Symbolique contribue significativement à l'amélioration de l'explication de la variabilité du TTR Grand total de sorte que le modèle « âge et NS-f Non-Symbolique » explique 82.3 % de la variance. Enfin, le score au NS-f Symbolique contribue significativement à l'amélioration de l'explication de la variabilité du TTR Grand total de sorte que le modèle « âge et NS-f Non-Symbolique et NS-f Symbolique » explique 80.4 % de la variance. Les résultats suggèrent une bonne prédictivité

du NS-f mesurant les habiletés de base à traiter le nombre et les numérosités sur le TTR mesurant les capacités arithmétiques.

**Tableau 5**

*Résumé de l'analyse de régression hiérarchique pour le TTR Grand total*

Variable	R	R <sup>2</sup>	R <sup>2</sup> ajusté	β	t
Étape 1					
Âge	.702	.493	.491	.702	16.041**
Étape 2	.823	.523	.520		
Âge				.650	14.652**
NS-f Non-Symbolique				.183	4.117**
Étape 2	.804	.647	.643		
Âge				.419	9.280**
NS-f Non-Symbolique				.017	0.393
NS-f Symbolique				.475	9.581**

\* $p < .05$ , \*\* $p < .01$

### 3. Analyse des performances au TTR et NS-f selon le niveau de classe, le genre, l'école et les conditions de passation

#### a. Relation entre les performances et le niveau de classe

Une série d'ANOVAs a été réalisée pour évaluer l'effet du niveau de classe sur les scores obtenus à chaque test. L'analyse sur le TTR a été réalisée pour les 1<sup>ère</sup> à 6<sup>ème</sup> années concernant l'addition, la soustraction et le Petit total ainsi que pour les 3<sup>ème</sup> à 6<sup>ème</sup> années concernant les autres scores. Un effet principal du niveau de classe attendu est observé pour chacun des sept scores du TTR (respectivement  $F(5,588) = 275.637$ ,  $p < .001$ ,  $\eta_p^2 = .701$  pour le score Addition,  $F(5,588) = 187.781$ ,  $p < .001$ ,  $\eta_p^2 = .615$  pour le score Soustraction,  $F(3,371) = 165.475$ ,  $p < .001$ ,  $\eta_p^2 = .572$  pour le score Multiplication,  $F(3,371) = 110.364$ ,  $p < .001$ ,  $\eta_p^2 = .472$  pour le score Division,  $F(3,371) = 106.830$ ,  $p < .001$ ,  $\eta_p^2 = .463$  pour le score Opérations mêlées,  $F(5,588) = 268.575$ ,  $p < .001$ ,  $\eta_p^2 = .695$ , pour le Petit total et  $F(3,371) = 145.461$ ,  $p < .001$ ,  $\eta_p^2 = .540$  pour le Grand total). Des analyses de comparaison ( $t$ -tests avec correction Bonferroni) concernant les comparaisons inter-niveaux montrent que toutes les différences sont significatives ( $p < .001$  ou  $p = .001$ ), excepté celle entre les 5<sup>ème</sup> et 6<sup>ème</sup> années pour la soustraction ( $p = .091$ ).

L'analyse sur le NS-f a été réalisée pour l'ensemble des sept niveaux de classe pour l'ensemble des deux écoles E1 et E2. Tout d'abord, un effet principal du niveau de classe attendu est observé pour chacun des trois scores du NS-f (respectivement  $F(6,581) = 65.286$ ,  $p < .001$ ,  $\eta_p^2 = .403$  pour le score Non-Symbolique,  $F(6,581) = 147.557$ ,  $p < .001$ ,  $\eta_p^2 = .604$  pour le score Symbolique et  $F(6,581) = 140.546$ ,  $p < .001$ ,  $\eta_p^2 = .592$  pour le score total). Les Tableaux 6, 7 et 8 indiquent les résultats des analyses de comparaison ( $t$ -tests avec correction Bonferroni) concernant, respectivement, le score Non-Symbolique, Symbolique et total. Les résultats suggèrent que le niveau de classe influence globalement les scores au TTR et au NS-f, justifiant ainsi de produire des normes différenciées pour chaque niveau.

**Tableau 6**

*Différences entre chaque niveau de classe pour le score Non-Symbolique du Numeracy Screener version française (NS-f)*

	1 <sup>e</sup> année	2 <sup>e</sup> année	3 <sup>e</sup> année	4 <sup>e</sup> année	5 <sup>e</sup> année	6 <sup>e</sup> année
Maternelle	$p = 1,000$	$p < 0,001$				
1 <sup>e</sup> année		$p < 0,001$				
2 <sup>e</sup> année			$p = 0,193$	$p < 0,001$	$p < 0,001$	$p < 0,001$
3 <sup>e</sup> année				$p < 0,001$	$p = 0,833$	$p < 0,001$
4 <sup>e</sup> année					$p = 0,262$	$p = 1,000$
5 <sup>e</sup> année						$p < 0,001$

Note. Les valeurs de  $p$  en gras indiquent des différences significatives.

**Tableau 7**

*Différences entre chaque niveau de classe pour le score Symbolique du Numeracy Screener version française (NS-f)*

	1 <sup>e</sup> année	2 <sup>e</sup> année	3 <sup>e</sup> année	4 <sup>e</sup> année	5 <sup>e</sup> année	6 <sup>e</sup> année
Maternelle	$p = 1,000$	$p < 0,001$				
1 <sup>e</sup> année		$p < 0,001$				
2 <sup>e</sup> année			$p = 0,042$	$p < 0,001$	$p < 0,001$	$p < 0,001$
3 <sup>e</sup> année				$p < 0,001$	$p < 0,001$	$p < 0,001$
4 <sup>e</sup> année					$p = 0,483$	$p < 0,001$
5 <sup>e</sup> année						$p = 0,001$

Note. Les valeurs de  $p$  en gras indiquent des différences significatives.

**Tableau 8**

*Différences entre chaque niveau de classe pour le score total du Numeracy Screener version française (NS-f)*

	1 <sup>e</sup> année	2 <sup>e</sup> année	3 <sup>e</sup> année	4 <sup>e</sup> année	5 <sup>e</sup> année	6 <sup>e</sup> année
Maternelle	$p = 1,000$	$p < 0,001$				
1 <sup>e</sup> année		$p < 0,001$				
2 <sup>e</sup> année			$p = 0,019$	$p < 0,001$	$p < 0,001$	$p < 0,001$
3 <sup>e</sup> année				$p < 0,001$	$p < 0,001$	$p < 0,001$
4 <sup>e</sup> année					$p = 1,000$	$p < 0,001$
5 <sup>e</sup> année						$p < 0,001$

Note. Les valeurs de  $p$  en gras indiquent des différences significatives.

## b. Relation entre les performances et le genre

Une série d'ANOVAs a été réalisée pour évaluer l'effet du genre et du niveau de classe sur les scores obtenus à chaque test. À noter que cette analyse n'est réalisée que pour l'école E2 car nous ne disposons pas des données de genre pour l'école E1. Un effet principal du niveau de classe attendu est observé pour chacun des sept scores du TTR (Addition, Soustraction, Multiplication, Division, Mêlé, Petit total, Grand total). Un effet principal du niveau de classe attendu est observé pour chacun des trois scores du NS-f (Non-Symbolique, Symbolique, Score total).

Concernant les scores du TTR, aucun effet principal de genre n'est observé pour le score en addition ( $F(1,239) = 3.374, p = .067$ ) et en multiplication ( $F(1,141) = 1.778, p = .185$ ). En revanche, un effet principal de genre est observé pour le score en soustraction ( $F(1,239) = 9.535, p = .002, \eta_p^2 = .038$ ), en division ( $F(1,141) = 6.470, p = .012, \eta_p^2 = .044$ ), en opérations mêlées ( $F(1,141) = 5.421, p = .021, \eta_p^2 = .037$ ), en Petit total ( $F(1,239) = 7.140, p = .008, \eta_p^2 = .029$ ) et en Grand total ( $F(1,141) = 6.894, p = .010, \eta_p^2 = .047$ ). Les garçons obtiennent des

performances significativement supérieures à celles des filles à chaque score. Aucune interaction Genre x Niveau de classe n'est observée pour les sept scores.

Aucun effet principal de genre n'est observé pour les scores du NS-f : aucun effet principal de genre ( $F(1,191) = 1.234, p = .268$ ) et aucune interaction Genre x Niveau de classe ( $F(3,191) = 0.870, p = .457$ ) concernant le score Non-Symbolique ; aucun effet principal de genre ( $F(1,191) = 1.714, p = .192$ ) et aucune interaction Genre x Niveau de classe ( $F(3,191) = 2.249, p = .084$ ) concernant le score Symbolique ; aucun effet principal de genre ( $F(1,191) = 1.954, p = .164$ ) et aucune interaction Genre x Niveau de classe ( $F(3,191) = 1.667, p = .175$ ) concernant le score total. Les résultats suggèrent donc que le genre influence différemment les scores au TTR et au NS-f, justifiant ainsi de produire des normes groupées pour les deux genres pour le NS-f et pour les scores en addition et en multiplication au TTR et différenciées pour les deux genres pour les autres scores du TTR.

### c. Performances selon les conditions de passation

Une série d'ANOVAs a été réalisée pour évaluer l'effet de forme du NS-f passée (forme A, forme B) sur les performances obtenues au NS-f (au score Non-Symbolique, au score Symbolique et au score total) en fonction du niveau de classe. Un effet principal du niveau de classe attendu est observé pour chacun des trois scores. Concernant le score Non-Symbolique, les résultats montrent un effet principal de la forme du NS-f ( $F(1,574) = 47.594, p < .001, \eta_p^2 = .077$ ), suggérant que les enfants ayant passé la forme A (partie symbolique puis partie non-symbolique) ont de meilleures performances que ceux ayant passé la forme B. L'interaction Forme x Niveau de classe est non significative. Concernant le score Symbolique, les résultats montrent un effet principal de la forme du NS-f ( $F(1,574) = 19.976, p < .001, \eta_p^2 = .034$ ) suggérant que les enfants ayant passé la forme B (partie non-symbolique puis partie symbolique) ont de meilleures performances que ceux ayant passé la forme A. L'interaction Forme x Niveau de classe est non significative. Concernant le score total du NS-f, aucun effet principal de la forme n'est observé pour ( $F(1,574) = 0.466, p = .495$ ). L'interaction Forme x Niveau de classe est non significative.

Une série d'ANOVAs a été réalisée pour évaluer l'effet de l'ordre de passation du TTR et du NS-f sur les performances obtenues à chacun des deux tests en fonction du niveau de classe. Un effet principal du niveau de classe attendu est observé pour chacun des deux tests. Concernant le NS-f, les résultats montrent un effet principal de l'ordre de passation sur le score total du NS-f ( $F(1,479) = 4.137, p = .043, \eta_p^2 = .009$ ), suggérant que les enfants ayant passé le TTR puis le NS-f ont de meilleures performances que ceux ayant passé les tests dans l'ordre inverse. Concernant le TTR, les résultats ne montrent aucun effet principal de l'ordre de passation sur le Petit total ( $F(1,479) = 1.198, p = .274$ ) et sur le Grand total ( $F(1,264) = 0.008, p = .929$ ). L'interaction Ordre x Niveau de classe est non significative. Les résultats suggèrent que l'ordre de passation influence différemment les scores au TTR et au NS-f. En raison de ces inconsistances, des normes indifférenciées quel que soit l'ordre de passation seront développées pour les deux tests.

### d. Comparaison des performances entre les deux écoles

Une série d'ANOVAs a été réalisée pour évaluer l'effet de l'école et du niveau de classe sur les performances obtenues au NS-f (score Non-Symbolique, score Symbolique et score total) pour les classes de maternelle, 1<sup>ère</sup>, 2<sup>ème</sup> et 3<sup>ème</sup> années. Un effet principal du niveau de classe attendu

est observé pour chacun des trois scores. Concernant le score Non-Symbolique, les résultats montrent un effet principal d'école ( $F(1,404) = 22.870, p < .001, \eta_p^2 = .054$ ) et une interaction École x Niveau ( $F(2,404) = 7.264, p < .001, \eta_p^2 = .051$ ). Des analyses de comparaison (*t*-tests avec correction Bonferroni) montrent que les enfants de maternelle et de 1<sup>ère</sup> année de E2 sont meilleurs que ceux de E1 ( $p < .001$ ) mais aucune différence n'est notée pour les enfants de 2<sup>ème</sup> et 3<sup>ème</sup> années. Concernant le score Symbolique, les résultats montrent un effet principal d'école ( $F(1,404) = 9.207, p = .003, \eta_p^2 = .022$ ) et une interaction École x Niveau ( $F(2, 404) = 16.368, p < .001, \eta_p^2 = .108$ ). Des analyses de comparaison (*t*-tests avec correction Bonferroni) montrent que les enfants de maternelle de E2 sont meilleurs que ceux de E1 ( $p < .001$ ) mais aucune différence n'est notée pour les enfants de 1<sup>ère</sup>, 2<sup>ème</sup> et 3<sup>ème</sup> années. Concernant le score total, les résultats montrent un effet principal d'école ( $F(1,404) = 19.018, p < .001, \eta_p^2 = .045$ ) et une interaction École x Niveau ( $F(2,404) = 15.393, p < .001, \eta_p^2 = .103$ ). Des analyses de comparaison (*t*-tests avec correction Bonferroni) montrent que les enfants de maternelle de E2 sont meilleurs que ceux de E1 ( $p < .001$ ) mais aucune différence n'est notée pour les enfants de 1<sup>ère</sup>, 2<sup>ème</sup> et 3<sup>ème</sup> années.

Une série d'ANOVAs a été réalisée pour évaluer l'effet de l'école et du niveau de classe sur les performances obtenues au TTR pour les classes de 1<sup>ère</sup> à 6<sup>ème</sup> année. Un effet principal du niveau de classe attendu est observé pour chacun des deux scores. Concernant le score Petit total, les résultats montrent un effet principal d'école ( $F(1,582) = 15.427, p < .001, \eta_p^2 = .026$ ) et une interaction École x Niveau ( $F(2,582) = 4.887, p < .001, \eta_p^2 = .040$ ). Des analyses de comparaison (*t*-tests avec correction Bonferroni) montrent que les enfants de 2<sup>ème</sup>, 3<sup>ème</sup> et 5<sup>ème</sup> années de E1 sont meilleurs que ceux de E2 mais aucune différence n'est notée pour les enfants de 1<sup>ère</sup>, 4<sup>ème</sup> et 6<sup>ème</sup> années. Concernant le score Grand total, les résultats montrent un effet principal d'école ( $F(1,367) = 13.044, p < .001, \eta_p^2 = .034$ ). Les enfants de E1 sont meilleurs que ceux de E2. Considérant que le niveau socioéconomique des deux écoles est proche, que le programme scolaire est identique (progression du MEES) et que les différences ne sont pas systématiquement et identiquement observées pour chaque score, une norme groupée sera ainsi privilégiée.

## E. En résumé.

À l'issue de ces analyses, il est décidé de créer une norme pour le TTR et pour le NS-f qui prendra en compte le niveau de classe et le genre (pour certains scores seulement), mais groupera les performances des deux écoles et des différentes conditions de passation.

## Normes

Considérant que la variable de genre influence différemment les différents scores obtenus au TTR, une norme différenciée ou commune aux filles et garçons est proposée par niveau scolaire pour chacun des sept scores. Le tableau 9 résume ainsi les normes de référence à considérer pour les scores Addition et Multiplication du TTR selon le niveau scolaire, alors que le tableau 10 résume ainsi les normes de référence à considérer pour les autres scores du TTR selon le niveau scolaire et le genre.

**Tableau 9**

*Normes de référence par niveau de classe pour les scores d'additions et de multiplications du Tempo Test Rekenen (TTR)*

		Effectif	Moyenne	ET	P5	P10	P25	P50	P75	P90	P95
<b>Addition</b>	1e	101	9	3	5	5	6	8	10	12	14
	2e	118	15	3	9	10	12	15	17	19	20
	3e	96	17	4	7,9	10,7	14,3	16,5	20	22	23
	4e	97	20	4	13	15	18	20	22	25,2	27,1
	5e	91	24	4	17	19,2	22	24	26	28	30
	6e	91	26	4	19,6	21	23	26	29	32,	34
<b>Multiplication</b>	3e	96	12	3,9	6,9	7,7	9	12	14	17,3	19
	4e	97	16,3	3,8	8,9	12	14	16	19	21	23
	5e	91	21,9	4,5	15	17	19	21	25	28	29
	6e	91	25,1	5,2	17	18,2	20	25	29	32,8	34

Notes. ET = écart-type ; P = percentile.

**Tableau 10**

*Normes de référence par niveau de classe et genre pour les scores de soustractions, divisions, opérations mêlées, petit total et grand total du Tempo Test Rekenen (TTR)*

		Effectif	Moyenne	ET	P5	P10	P25	P50	P75	P90	P95	
<b>Soustraction</b>	1e	Filles	17	6,6	3,2	2	2	5	6	8	10	15
		Garçons	26	6,3	3,6	1	2	2	7	9	10	12
	2e	Filles	32	8,8	3,3	3	4	7	8,5	10,5	13	15
		Garçons	27	10,1	2,9	6	6	7	10	13	14	14
	3e	Filles	23	11,9	3,9	6	8	8	12	13	16	19
		Garçons	23	14,9	4,6	8	9	11	16	18	19	21
	4e	Filles	20	16	3,6	10,5	12	13,5	16	17	21,5	23
		Garçons	26	17,8	5,2	9	12	15	16,5	23	26	26
	5e	Filles	11	17,5	4	13	13	14	17	19	24	25
		Garçons	15	19,8	6,6	7	10	16	19	25	29	30
	6e	Filles	14	21,1	5,8	7	14	17	23	25	27	28
		Garçons	22	23	4,4	17	17	19	24	26	28	29
<b>Division</b>	3e	Filles	23	7,1	3,9	2	3	4	7	12	12	13
		Garçons	23	9	2,9	5	6	7	9	11	12	13
	4e	Filles	20	7,9	3,7	1,5	3	6	7	10	13,5	15,5
		Garçons	26	10	5	2	3	7	9,5	16	16	16
	5e	Filles	11	10,1	4,3	5	6	7	9	14	15	19
		Garçons	15	12,2	4,8	4	7	7	14	16	19	19
6e	Filles	14	15,3	4,8	6	7	13	15,5	19	21	22	
	Garçons	22	16,2	6,6	8	9	11	15,5	21	23	29	
<b>Mêlé</b>	3e	Filles	23	9,4	3,2	5	5	7	9	11	12	16
		Garçons	23	11,5	3,6	6	8	9	11	14	15	17
	4e	Filles	20	12,6	4,2	5,5	6,5	9,5	13	15	18	19
		Garçons	26	15,6	4,3	8	10	13	15,5	19	21	22
	5e	Filles	11	16,4	3,5	10	13	13	16	20	20	20
		Garçons	15	18,5	5,8	8	12	15	17	25	27	27
6e	Filles	14	20,6	5,8	10	14	16	21	25	27	31	
	Garçons	22	21,1	6	15	15	18	19,5	25	30	31	
<b>Petit total</b>	1e	Filles	17	16,2	4,8	10	10	13	15	18	23	28
		Garçons	26	15,6	6,4	6	6	12	16	20	24	25
	2e	Filles	32	22,2	6,4	11	15	19	21,5	26,5	31	32
		Garçons	27	23,3	5	16	16	20	24	28	30	31
	3e	Filles	23	26,5	6,8	15	19	23	26	30	33	39
		Garçons	23	32,1	8,9	19	23	26	33	38	40	47
	4e	Filles	20	34,7	6,3	25,5	26,5	31	34	38,5	41	46,5
		Garçons	26	38,4	9,2	23	29	33	36,5	46	53	54
	5e	Filles	11	39,9	6,3	34	35	35	37	45	50	52

	Garçons	15	43,3	13,2	11	27	36	43	54	61	62	
6e	Filles	14	47,1	9,4	25	34	42	49,5	53	58	59	
	Garçons	22	50	8,5	37	38	43	50	56	60	64	
<b>Grand total</b>	3e	Filles	23	53,3	15,3	29	37	44	52	61	75	78
	Garçons	23	64,3	16	34	53	55	66	74	75	94	
4e	Filles	20	69,3	15,7	45	47	63,5	69,5	77,5	89	97,5	
	Garçons	26	79,8	21	49	49	66	77,5	96	110	111	
5e	Filles	11	86,9	15,6	67	75	75	82	100	101	119	
	Garçons	15	95,8	26,4	39	63	85	94	119	133	136	
6e	Filles	14	106,1	21,1	60	78	95	107,5	118	127	143	
	Garçons	22	109,6	23,3	81	82	93	104,5	126	143	146	

Notes. ET = écart-type ; P = percentile.

Considérant que la variable de genre n'influence pas la performance au NS-f, une norme commune aux filles et aux garçons est proposée par niveau scolaire pour chacun des trois scores. Le tableau 11 résume ainsi les normes de référence à considérer pour le NS-f.

**Tableau 11**

*Normes de référence par niveau de classe pour les scores du Numeracy Screener version française (NS-f)*

		Effectif	Moyenne	ET	P5	P10	P25	P50	P75	P90	P95
<b>NS-f Non-Symbolique</b>	Maternelle	97	29,4	9,1	14	15	24	28	35	39,2	47
	1e	101	29,9	7,2	15	20	28	28	34	38,8	42
	2e	118	36,7	7,0	28	28	34	36	40	46,1	52
	3e	96	39,6	7,2	28	29	35	38,5	44	50	52
	4e	52	46	8,3	36	37	41,5	45	52	56,8	57,9
	5e	67	42,3	9,7	28	33,8	36	41	49	55,2	61,1
	6e	57	48,8	8,4	36	39	45	49	54	57,3	62,4
<b>NS-f Symbolique</b>	Maternelle	97	30,5	13,0	6	14,8	22	31	39	48,2	53
	1e	101	32,3	7,6	20	23	28	33	37	42	44
	2e	118	41,6	8,7	28	28,9	35	42	49	55,1	56
	3e	96	45,7	9,8	31	33	39	45	56	56	62,2
	4e	52	55,5	6,0	42	46	51,5	56	60	61,9	65,9
	5e	67	59,6	11,0	45	48	53	56,9	64,6	72,3	78,1
	6e	57	66,7	10,2	52	52	58,9	67,2	74,7	80	82
<b>NS-f Total</b>	Maternelle	97	60	20,0	28	34	46	60	73	87	93
	1e	101	62,2	13,2	42	44,4	54	62	70	79	84
	2e	118	78,3	13,5	59	61,9	70	79	87	95	102
	3e	96	85,4	14,8	64	66	74,5	85	97	104,3	112
	4e	52	101,5	11,2	85	86,3	94	101	110,4	114,8	119
	5e	67	101,9	17,8	80	82	92	98,9	110	126,2	131,1
	6e	57	115,6	14,9	91	95,8	105,9	115,3	127,2	131,9	138,7

Notes. ET = écart-type ; P = percentile.

## ----- DISCUSSION -----

L'objectif de la présente étude était d'évaluer les qualités psychométriques de deux tests d'évaluation mathématique, le TTR et le NS-f, dans le but ultime de développer des normes franco-québécoises qui seront utiles au dépistage et à l'identification des enfants à risque de présenter un trouble des apprentissages en mathématiques. Pour cela, un total de 695 enfants a été testé et leurs résultats analysés.

### 1. Comparaison avec d'autres périodes temporelles et régions géographiques

Tout d'abord, les résultats ont montré que les performances de deux groupes de 3<sup>e</sup> année de primaire franco-québécois évalués à deux temps différents ne différaient pas aux scores du TTR, suggérant une fidélité temporelle du TTR. Ces résultats nécessiteraient évidemment d'être reproduits avec un même groupe à deux temps différents et sur l'ensemble des niveaux scolaires, de même que pour le NS-f.

Cependant, les résultats ont montré une unique différence entre les enfants franco-québécois de 3<sup>e</sup> année en 2014 et ceux en 2018 en divisions au TTR. De plus, des différences ont été identifiées entre les enfants franco-québécois de la présente étude et les enfants canadiens anglophones d'Ontario et les enfants français d'autre part. En effet, les résultats ont montré qu'au NS-f Non-Symbolique, les 1<sup>ère</sup> année d'Ontario étaient meilleurs que ceux du Québec alors qu'aucune différence n'est observée pour les 2<sup>ème</sup> et 3<sup>ème</sup> années. Au NS-f Symbolique, le pattern de réussite est inversé : aucune différence n'est observée pour les 1<sup>ère</sup> année, alors que les 2<sup>ème</sup> et 3<sup>ème</sup> années du Québec étaient meilleurs que ceux d'Ontario. Une analyse a montré que, pour l'ensemble des trois scores au NS-f, les jeunes français de 5<sup>ème</sup> et 6<sup>ème</sup> année obtenaient des performances inférieures à celles des jeunes franco-québécois. Les résultats suggèrent donc que la fidélité géographique du NS-f n'est pas toujours respectée. Cela pourrait s'expliquer par de multiples facteurs (programme scolaire, langue d'apprentissage des mathématiques, etc.) mais ces hypothèses nécessiteraient d'être investiguées de manière plus systématique et approfondie. Ces différences entre provinces et pays justifient donc tout à fait de produire des normes spécifiques au contexte franco-québécois.

### 2. Validité des tests TTR et NS-f

Plusieurs analyses ont été réalisées et ont mis en évidence plusieurs types de validité. Tout d'abord, les résultats ont montré un fort coefficient  $\alpha$  de Cronbach et des corrélations fortes entre les scores des deux subtests du NS-f suggérant ainsi une bonne cohérence interne du NS-f. De même, un fort coefficient  $\alpha$  de Cronbach et des corrélations fortes entre les scores des cinq subtests du TTR entre eux, ainsi qu'avec les scores totaux, ont été identifiés, suggérant ainsi une bonne cohérence interne du TTR.

De plus, des corrélations moyennes à fortes ont été identifiées entre les scores du NS-f et les scores du TTR. Une analyse de régression a montré que les habiletés arithmétiques (mesurées par le TTR) étaient prédites, à plus de 70 %, par l'âge et les habiletés de traitement des numérosités (mesurées par le NS-f Non-Symbolique) et des nombres arabes (mesurées par le NS-f Symbolique). L'ensemble de ces résultats démontre ainsi une bonne prédictivité du NS-f mesurant les habiletés de base à traiter le nombre et les numérosités sur le TTR mesurant les capacités arithmétiques et, *de facto*, une bonne cohérence interne de notre dispositif global de

dépistage des risques de présenter un trouble des apprentissages en mathématiques (TTR et NS-f).

### 3. Comparaisons selon le niveau de classe, le genre, l'école et les conditions de passation

Des analyses comparatives ont mis en évidence que tous les scores des subtests (TTR et NS-f) montraient des différences significatives inter-niveaux. Toutefois, bien que le subtest Non-Symbolique du NS-f tende à montrer un effet plafond (les 3<sup>ème</sup> année ne différaient pas des 4<sup>ème</sup> année, qui eux-mêmes ne différaient pas des 5<sup>ème</sup> année par exemple), on observe tout de même une légère amélioration des performances de sorte que, par exemple, les 3<sup>ème</sup> années et les 5<sup>ème</sup> années différaient bien des 6<sup>ème</sup> années qui obtenaient de meilleures performances. Ce subtest vise à évaluer les habiletés de base du traitement des numérosités. Ces habiletés sont considérées comme quasi-innées car elles sont présentes chez le bébé (par exemple, Starr, Libertus & Brannon, 2013), le tout jeune enfant (par exemple, Halberda, Mazzocco & Feigenson, 2008) et l'adulte sans langage et culture numériques (par exemple, Frank, Everett, Fedorenko & Gibson, 2008). Il semble alors tout à fait attendu que les performances soient déjà très bonnes chez le tout jeune scolarisé en maternelle et n'évoluent guère ensuite. En revanche, tous les autres scores ont montré une amélioration progressive, de niveau en niveau, tant pour les habiletés de base à traiter le nombre arabe (NS-f Symbolique) que pour les habiletés de fluence arithmétique (TTR) : plus les enfants augmentent en niveau de classe, plus ils gagnent en efficacité (précision et vitesse) de traitement du nombre arabe et des capacités arithmétiques. Ces résultats appuient ainsi la validité de construit du NS-f et du TTR en démontrant que les performances évoluent ou non en fonction du niveau de classe selon l'habileté cible, de sorte qu'une norme par niveau est fournie.

De plus, les analyses montrent que le genre influence différemment les deux subtests. Tout d'abord, le genre n'a aucun effet sur les performances au NS-f suggérant ainsi que les habiletés de traitement des numérosités et du nombre ne sont pas différentes chez les filles et les garçons. Ce résultat est en accord avec les études précédemment menées directement sur ces capacités. Par exemple, Halberda et al. (2008) ne montraient aucune différence significative entre les filles et les garçons à une tâche de comparaison de numérosités. De même, tout récemment, Kersey, Braham, Csumitta, Libertus et Cantlon (2018) ont évalué 500 enfants âgés de 6 mois à 8 ans et n'ont montré aucune différence entre les filles et les garçons sur des tâches telles que la perception de numérosité suggérant donc que les jeunes filles et les jeunes garçons sont équipés de manière égale pour les mathématiques. Spelke (2005) soutient d'ailleurs l'idée que les filles et les garçons ont des capacités numériques primaires égales et qu'aucun argument ne saurait montrer, à ce jour, que filles et garçons diffèrent biologiquement. En revanche, les résultats de la présente étude ont mis en évidence que les garçons montrent généralement de meilleurs résultats que les filles pour la fluence arithmétique. Ces résultats sont par exemple en accord avec l'étude de Carr et Davis (2001) qui montraient que les filles et les garçons différaient en termes de réponses correctes mais contredisent Wei et al. (2012) qui montraient une supériorité féminine. Aucun consensus ne semble se dégager des études concernant l'effet du genre sur la fluence arithmétique.

Il est également surprenant d'observer des différences entre les filles et les garçons en arithmétique alors que le genre n'influence pas les habiletés de base de traitement des numérosités et du nombre, qui sont elles-mêmes prédictives des capacités arithmétiques. Plusieurs hypothèses peuvent être avancées. Premièrement, on pourrait supposer que les filles

et les garçons diffèrent en fluence arithmétique car cette activité nécessite des compétences cognitives pour lesquelles une supériorité serait notée chez les garçons telles que les capacités visuospatiales ou la récupération en mémoire à long terme (voir la méta-analyse de Zhu, 2007 pour une revue). Toutefois, Spelke (2005) et Flores-Mendoza, Ardila, Rosas, Lucio, Gallegos et al. (2018) démontrent qu'il n'y a à ce jour aucun appui consensuel pour affirmer que les garçons ont un profil d'habiletés numériques, spatiales, cognitives menant à de meilleures performances mathématiques.

Une seconde hypothèse, plus plausible, a trait davantage à des facteurs environnementaux tels que l'existence de forts stéréotypes sociaux présents dans la société (Frost, Hyde & Fennema, 1994 ; Huguet & Régner, 2007, 2009), une anxiété mathématique plus grande et une moins forte confiance en leurs capacités mathématiques chez les filles (Maloney, Waechter, Risko & Fugelsang, 2012 ; Ramirez, Shaw & Maloney, 2018 ; étude PISA - Programme International pour le Suivi des Acquis des élèves – 2009, <http://www.oecd.org/pisa/46624382.pdf>), des croyances et des attitudes des enseignants qui sont parfois différentes selon le genre des élèves (Li, 1999). D'ailleurs, Entwisle et Baker (1983) montraient par exemple que les garçons développaient de plus grands espoirs quant à leur capacité arithmétique, espoirs eux-mêmes liés aux attentes parentales. Zhu (2007) conclut d'ailleurs qu'une combinaison de facteurs biologiques, psychologiques, environnementaux (éducation et expérience) expliquerait les différences de performances mathématiques liées au genre. Ces résultats appuient ainsi la validité de construit du NS-f et du TTR en démontrant que les performances divergent ou non en fonction du genre selon l'habileté cible. Une norme groupée est fournie pour le NS-f alors qu'une norme distincte pour les filles et les garçons est fournie pour le TTR.

Des analyses ont été réalisées pour investiguer l'effet de la forme et de l'ordre de passation des deux tests. Tout d'abord, les résultats ont montré que les enfants ayant passé la forme A (partie symbolique puis partie non-symbolique) ont de meilleures performances au NS-f Non-Symbolique que ceux ayant passé la forme B, alors que le pattern inverse est observé pour le NS-f Symbolique. Cette observation est similaire à celle faite par Michaud (2018). Ces résultats pourraient s'expliquer par deux raisons. La première serait que la comparaison symbolique et la comparaison non-symbolique entretiennent des relations étroites (Hyde, Berteletti & Mou, 2016), de sorte que travailler l'une améliore l'autre. La seconde serait que le deuxième subtest est toujours mieux réussi que le premier car les enfants ont mieux compris et sont plus à l'aise avec la consigne et la procédure. Notons tout de même que peu importe l'ordre, la performance totale ne diffère pas. De même, les performances sont les mêmes au TTR et au NS-f peu importe l'ordre de passation des deux tests. Pour cette raison, une norme indifférenciée a été réalisée.

Enfin, une comparaison entre nos deux écoles franco-québécoises a montré qu'au NS-f, les enfants des deux écoles ne diffèrent pas ou les enfants de E2 sont meilleurs que ceux de E1 (selon subtest et niveau de classe). De manière surprenante, le pattern est inversé pour le TTR : les enfants des deux écoles ne diffèrent pas ou les enfants de E1 sont meilleurs que ceux de E2 (selon subtest et niveau de classe). Considérant que le niveau socioéconomique des deux écoles est proche (décile 7 pour E1 et décile 8 pour E2 selon les indices de défavorisation du MEES), il est difficilement imaginable que ce facteur puisse expliquer les différences. En revanche, bien que le programme scolaire soit identique (progression du MEES), les élèves de 2<sup>ème</sup> et 3<sup>ème</sup> cycle de l'école E1 ont participé au projet « CalculoManiac » visant à développer les stratégies de calcul mental (d'autant plus que l'école E1 a plus d'enfants en 5<sup>ème</sup> et 6<sup>ème</sup> années qu'E2) et cela pourrait tout à fait expliquer pourquoi les jeunes de certains niveaux de classe de E1 obtiennent des performances supérieures aux jeunes de E2 au test de fluence arithmétique (cela

en dépit de leurs moindres habiletés de base de traitement du nombre et des numérosités). Considérant que le niveau socioéconomique des deux écoles est proche, que le programme scolaire est identique (progression du MEES) et que les différences ne sont pas systématiquement et identiquement observées pour chaque score, une norme groupée a été privilégiée.

## ----- CONCLUSION -----

En conclusion, les résultats de l'étude suggèrent que le présent dispositif de dépistage comportant le NS-f et le TTR a une bonne validité. Les tests NS-f (initialement version canadienne anglophone NS) et TTR (initialement version néerlandaise) visent à évaluer, respectivement, les habiletés de base à comparer les nombres et les numérosités et les habiletés de fluence arithmétique. Ils présentent une bonne cohérence interne, une bonne validité prédictive et une bonne validité de construit. Considérant l'ensemble de ces résultats, des normes ont été établies en regard des performances des jeunes scolarisés de la maternelle à la 6<sup>ème</sup> année de primaire pour le NS-f et de la 1<sup>ère</sup> à la 6<sup>ème</sup> année de primaire pour le TTR, soit par niveau de classe soit par niveau de classe et de genre.

En termes d'implications cliniques, ces résultats permettent de fournir des données importantes pour les personnes en charge du dépistage et de l'identification des enfants à risque de présenter un trouble des apprentissages en mathématiques dans le contexte franco-québécois de la maternelle à la 6<sup>ème</sup> année de primaire. Considérant que les enfants ont été testés pendant le mois de mai, un professionnel évaluant un enfant de 4<sup>ème</sup> année de primaire en début d'année (septembre) devrait donc comparer les scores de cet enfant à la norme correspondant aux jeunes de 3<sup>ème</sup> année (testé en fin d'année scolaire, mai). Puisqu'il s'agit de dépistage – et non de diagnostic – il semble adéquat de considérer qu'un enfant dont les scores se situent sous le percentile 25 au NS-f et/ou au TTR est en difficulté, puisse être à risque de présenter un trouble des apprentissages en mathématiques et mériterait donc toute l'attention de l'équipe scolaire, enseignants et professionnels en charge d'apporter une aide supplémentaire. Enfin, ce travail soulève plusieurs perspectives telles que poursuivre l'évaluation de la validité et de la fidélité de ces deux outils telles que la validité concomitante, la validité discriminante, la sensibilité et la spécificité, la fidélité temporelle. Ce travail aboutira à consolider les connaissances et les normes pour les enfants franco-québécois.

## ----- REMERCIEMENTS -----

Nous remercions Audrey Cormier et Élodie Lemay, orthopédagogues, ainsi que Caroline Bisson, conseillère pédagogique en mathématiques et en sciences au primaire, pour leur aide précieuse dans le recrutement des écoles participantes et la passation des tests. Nous remercions également les directrices d'établissement, les enseignant(e)s et les enfants pour leur participation.

## ----- BIBLIOGRAPHIE -----

- American Psychiatric Association (2016). *Mini DSM-5. Critères diagnostiques*. Issy-les-Moulineaux : Elsevier Masson.
- American Psychiatric Association (2013). *Diagnostic and Statistical manual of Mental disorders (DSM-5)*. Arlington, VA: APA.
- Bynner, J., & Parsons, S. (1997). *Does Numeracy Matter? Evidence from the National Child Development Study on the Impact of Poor Numeracy on Adult Life*. England, UK: Basic Skills Agency. Consulté le 12.02.2020 de ERIC : <https://files.eric.ed.gov/fulltext/ED406585.pdf>
- Butterworth, B. (2005). Developmental Dyscalculia. In J.I.D. Campbell (Ed), *Handbook of Mathematical Cognition* (pp. 455-467). Hove and New-York : Psychology Press.
- Butterworth, B. (1999). *The Mathematical Brain*. London, UK: MacMillan.
- Campbell, J. I. D., & Oliphant, M. (1992). Representation and Retrieval of Arithmetic Facts: A Network-Interference Model and Simulation. *Advances in Psychology*, 91, 331-364. [https://doi.org/10.1016/S0166-4115\(08\)60891-2](https://doi.org/10.1016/S0166-4115(08)60891-2)
- Carr, M., & Davis, H. (2001). Gender Differences in Arithmetic Strategy Use: A Function of Skill and Preference. *Contemporary Educational Psychology*, 26(3), 330-347. <https://doi.org/10.1006/ceps.2000.1059>
- Dehaene, S. (2010). *La bosse des maths, 15 ans après*. Paris: Odile Jacob.
- Dellatolas, G., & von Aster, M. (2006). *Zareki-R, Batterie pour l'évaluation du traitement des nombres et du calcul chez l'enfant*. Paris, France : ECPA.
- De Vos, T. (1992). *Tempo Test Rekenen (TTR). Arithmetic Number Fact Test*. Nijmegen: Berkhout.
- Dirks, E., Spyer, G., van Lieshout, E. C. D. M., & de Sonnevile, L. (2008). Prevalence of Combined Reading and Arithmetic Disabilities. *Journal of Learning Disabilities*, 41(5), 460–473. <https://doi.org/10.1177/0022219408321128>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, 43(6), 1428-1446. <https://doi.org/10.1037/0012-1649.43.6.1428> Erratum in *Developmental Psychology*, 44(1), 232. <https://doi.org/10.1037/0012-1649.44.1.217>
- Entwisle, D. R., & Baker, D. P. (1983). Gender and Young Children's Expectations for Performance in Arithmetic. *Developmental Psychology*, 19(2), 200-209. <https://doi.org/10.1037/0012-1649.19.2.200>
- Geary, D. C. (2011). Cognitive Predictors of Achievement Growth in Mathematics: A 5-Year Longitudinal Study. *Developmental Psychology*, 47(6), 1539-1552. PMC3210883 accès ouvert.

- Green, K.B., Gallagher, P.A. (2014). Mathematics for Young Children: A Review of the Literature with Implications for Children with Disabilities. *Başkent University Journal of Education*, 1(1), 81-92. Consulté le 12.02.2020 de [semanticscholar.org](https://pdfs.semanticscholar.org/af81/b01a544d518b68244626d91b470cd6b2a2ac.pdf) : <https://pdfs.semanticscholar.org/af81/b01a544d518b68244626d91b470cd6b2a2ac.pdf>
- Flores-Mendoza, C., Ardila, R., Rosas, R., Lucio, M.E., Gallegos, M., & Colareta, N.R. (2018). Intelligence, Latin America, and Human Capital. In *Intelligence Measurement and School Performance in Latin America* (pp.79-112). Springer, Cham.
- Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a Cognitive Technology: Evidence from Pirahã Language and Cognition. *Cognition*, 108(3), 819-824. Consulté le 12.02.2020 de University of California San Diego : <http://lhc.ucsd.edu/mca/Mail/xmcamail.2014-12.dir/pdf2Yb7JAO0ZG.pdf>
- Frost, L. A., Hyde, J. S., & Fennema, E. (1994). Gender, Mathematics Performance, and Mathematics-Related Attitudes and Affect: A Meta-Analytic Synthesis. *International Journal of Educational Research*, 21(4), 373-385. [https://doi.org/10.1016/S0883-0355\(06\)80026-1](https://doi.org/10.1016/S0883-0355(06)80026-1)
- Halberda, J., Mazocco, M. M., & Feigenson, L. (2008). Individual Differences in Non-Verbal Number Acuity Correlate with Maths Achievement. *Nature*, 455(7213), 665-668. <https://doi.org/10.1038/nature07246>
- Huguet, P., & Regner, I. (2007). Stereotype Threat among Schoolgirls in Quasi-Ordinary Classroom Circumstances. *Journal of Educational Psychology*, 99(3), 545-560. Consulté le 12.02.2020 de HAL : <https://hal.archives-ouvertes.fr/hal-01440632>
- Huguet, P., & Régner, I. (2009). Counter-Stereotypic Beliefs in Math Do not Protect School Girls from Stereotype Threat. *Journal of Experimental Social Psychology*, 45(4), 1024-1027. Consulté le 12.02.2020 de HAL : <https://hal.archives-ouvertes.fr/hal-00694723>
- Hyde, D. C., Berteletti, I., & Mou, Y. (2016). Approximate Numerical Abilities and Mathematics: Insight from Correlational and Experimental Training Studies. *Progress in Brain Research*, 227, 335-351. <https://doi.org/10.1016/bs.pbr.2016.04.011>
- Kersey, A. J., Braham, E. J., Csumitta, K. D., Libertus, M. E., & Cantlon, J. F. (2018). No Intrinsic Gender Differences in Children's Earliest Numerical Abilities. *Npj Science of Learning*, 3(1), 12. <https://doi.org/10.1038/s41539-018-0028-7> accès ouvert
- Lafay, A., Archambault, S., Vigneron, M., & Nosworthy, N. (2018). Version française du test Numeracy Screener (NS-f), un outil de dépistage des difficultés de traitement du nombre et des quantités. *Glossa*, 123,18-32. Consulté le 12.02.2020 de Andrews University : <https://digitalcommons.andrews.edu/pubs/1056/>
- Lafay, A., & Cattini, J. (2018). Analyse psychométrique d'outils d'évaluation mathématiques utilisés auprès des enfants francophones. *Revue Canadienne d'Orthophonie et d'Audiologie*, 42(2), 147-164. Consulté le 12.02.2020 de RCOA : <https://www.cjslpa.ca/detail.php?ID=1232&lang=fr>

Lafay, A., Macoir, J., & St-Pierre, M. C. (2018). Impairment of Arabic-and Spoken-Number Processing in Children with Mathematical Learning Disability. *Journal of Numerical Cognition*, 3(3), 620-641. Consulté le 12.02.2020 de semanticscholar.org : <https://pdfs.semanticscholar.org/e50c/d49fb9bfa237488e2359b7b24b2106d439b5.pdf>

Lafay, A., St-Pierre, M. C., & Macoir, J. (2019). Impairment of Non-Symbolic Number Processing in Children with Mathematical Learning Disability. *Journal of Numerical Cognition*, 5(1), 86-104. <https://doi.org/10.5964/jnc.v5i1.177> accès ouvert

Lafay, A., St-Pierre, M. C., & Macoir, J. (2017). The Mental Number Line in Dyscalculia: Impaired Number Sense or Access from Symbolic Numbers? *Journal of Learning Disabilities*, 50(6), 672-683. <https://doi.org/10.1177/0022219416640783>

Lafay, A., St-Pierre, M. C., & Macoir, J. (2015). Validation franco-québécoise du Tempo Test Rekenen pour l'évaluation des habiletés mathématiques auprès d'enfants de 8-9 ans. *Glossa*, 118, 27-39. Consulté le 12.02.2020 de docplayer : <https://docplayer.fr/20755920-Validation-franco-quebecoise-du-tempo-test-rekenen-pour-l-evaluation-des-habiletés-mathématiques-aupres-d-enfants-de-8-9-ans.html>

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. <http://glossa.fr/userdata/9cc1fa312abcc3924d0b3b9616f2318a.pdf>

Li, Q. (1999). Teachers' Beliefs and Gender Differences in Mathematics: A Review. *Educational Research*, 41(1), 63-76. <https://doi.org/10.1080/0013188990410106>

Maloney, E. A., Waechter, S., Risko, E. F., & Fugelsang, J. A. (2012). Reducing the Sex Difference in Math Anxiety: The Role of Spatial Processing Ability. *Learning and Individual Differences*, 22(3), 380-384. <https://doi.org/10.1016/j.lindif.2012.01.001>

Michaud, S. (2018). *Étalonnage du test Numeracy Screener chez une population française scolarisée en CM2 et 6<sup>ème</sup>*. Mémoire pour l'obtention du Certificat de Capacité d'Orthophoniste. Nice : Département Universitaire d'Orthophonie, Nice Sophia Antipolis.

Ministère de l'Éducation et de l'Enseignement Supérieur - MEES (2016-2017). *Indices de défavorisation*. Consulté le 12.02.2020 de : <http://www.education.gouv.qc.ca/references/indicateurs-et-statistiques/indices-de-defavorisation/>

Ministère de l'Éducation, de l'Enseignement supérieur et de la Recherche (2015). Cadre de référence et guide à l'intention du milieu scolaire – L'intervention auprès des élèves ayant des difficultés de comportement. Consulté le 12.02.2020 de : [http://www.education.gouv.qc.ca/fileadmin/site\\_web/documents/dpse/adaptation\\_serv\\_compl/14\\_00479\\_cadre\\_intervention\\_eleves\\_difficultés\\_comportement.pdf](http://www.education.gouv.qc.ca/fileadmin/site_web/documents/dpse/adaptation_serv_compl/14_00479_cadre_intervention_eleves_difficultés_comportement.pdf)

Ministère de l'Éducation, du Loisir et du Sport – MELS (2011). *Référentiel d'intervention en lecture pour les élèves de 10 à 15 ans*. Consulté le 12.02.2020 de : [http://www.education.gouv.qc.ca/fileadmin/site\\_web/documents/dpse/adaptation\\_serv\\_compl/Referentiel-Lecture\\_section2.pdf](http://www.education.gouv.qc.ca/fileadmin/site_web/documents/dpse/adaptation_serv_compl/Referentiel-Lecture_section2.pdf)

National Center on Response to Intervention - NCRTI (2010). *Essentials Components of RTI - A Closer Look at Response to Intervention*. Washington, DC: American Institutes for Research. Consulté le 12.02.2020 de <https://rti4success.org/resource/essential-components-rti-closer-look-response-intervention>

Noël, M.P., Rousselle, L. (2011). Developmental Changes in the Profiles of Dyscalculia: An Explanation Based on a Double Exact-and-Approximate Number Representation Model. *Frontiers in Human Neuroscience*, 5(165), 29-32. <https://doi.org/10.3389/fnhum.2011.00165> accès ouvert

Nosworthy, N. (2013). *An Investigation of the Association between Arithmetic Achievement and Symbolic and Nonsymbolic Magnitude Processing in 5-9 Year-Old Children: Evidence from a Paper-and-Pencil Test*. Dissertation, Western University, London, Ontario, Canada. <https://www.semanticscholar.org/paper/An-Investigation-of-the-Association-Between-and-and-Nosworthy/37f111c55b3f1a574ef614f98044e4d551e752d0>

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & 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. <https://doi.org/10.1371/journal.pone.0067918> accès ouvert

Nosworthy, N., Zheng, S., & Ansari, D. (2014). *Kindergarten children's number comparison skills predict later math scores: Evidence from a two-minute test*. Poster présenté au Sixth Annual Celebration of Research and Creative Scholarship, 2014, Andrews University, Michigan.

Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math Anxiety: Past Research, Promising Interventions, and a New Interpretation Framework. *Educational Psychologist*, 53(3), 145-164. <https://doi.org/10.1080/00461520.2018.1447384>

Ritchie, S. J., & Bates, T. C. (2013). Enduring Links from Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status. *Psychological Science*, 24(7), 1301-1308. <https://doi.org/10.1177/0956797612466268>

Spelke, E. S. (2005). Sex Differences in Intrinsic Aptitude for Mathematics and Science?: A Critical Review. *American Psychologist*, 60(9), 950-958. <https://doi.org/10.1037/0003-066X.60.9.950>

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Infants Show Ratio-dependent Number Discrimination Regardless of Set Size. *Infancy*, 18(6), 927-941. <https://doi.org/10.1111/infa.12008>

Thevenot, C. (2017). La dyscalculie développementale vue comme un déficit d'automatisation des procédures de comptage. *Rééducation Orthophonique*, 269, 113-123.

Von Aster, M.G., Shalev, R.S. (2007). Number Development and Developmental Dyscalculia. *Developmental Medicine and Child Neurology*, 49(11), 868-873. <https://doi.org/10.1111/j.1469-8749.2007.00868.x>

Wechsler, D. (2005, 2<sup>nd</sup> ed.). *Wechsler Individual Achievement Test (WIAT II)*. London, UK: The Psychological Corporation.

Wei, W., Lu, H., Zhao, H., Chen, C., Dong, Q., & Zhou, X. (2012). Gender Differences in Children's Arithmetic Performance Are Accounted for by Gender Differences in Language Abilities. *Psychological Science*, 23(3), 320-330. <https://doi.org/10.1177/0956797611427168>

Wilson, A.J., Dehaene, S. (2007). Number Sense and Developmental Dyscalculia. In D. Coch, G. Dawson, K.W. Fischer (Eds), *Human Behavior, learning, and the developing brain: atypical development* (pp. 212–238). New-York: Guilford Press. Consulté le 12.02.2020 de UNICOG : [http://www.unicog.org/publications/WilsonDehaene\\_inPress\\_Final\\_HBDB.pdf](http://www.unicog.org/publications/WilsonDehaene_inPress_Final_HBDB.pdf)

Zhu, Z. (2007). Gender Differences in Mathematical Problem Solving Patterns: A Review of Literature. *International Education Journal*, 8(2), 187-203. Consulté le 12.02.2020 de ERIC : <https://eric.ed.gov/?id=EJ834219>