INTERNATIONAL JOURNAL OF EARLY YEARS EDUCATION
2026, VOL. 34, NO. 2, 254–276
https://doi.org/10.1080/09669760.2026.2645342

Routledge
Taylor & Francis Group

The Early NUmerical DIdactical Test (ENUDIT): a French-language monitoring test to assess early numerical abilities of children from 3 to 6

Marie-Line Gardes [a], Marie-Caroline Croset [b], Adeline Samson[c] and Jérôme Prado [d]

[a]Haute École Pédagogique du Canton de Vaud, UER MS, Lausanne, Switzerland; [b]Université Grenoble Alpes, CNRS, LIG, Grenoble, France; [c]Université Grenoble Alpes, CNRS, LJK, Grenoble, France; [d]Centre de Recherche en Neurosciences de Lyon (CRNL), INSERM U1028 - CNRS UMR5292, Université de Lyon, Lyon, France

ABSTRACT

There are large individual differences in early numerical skills. Because these differences predict the growth of future math skills, it is important to monitor the development of early numerical abilities in children. However, there is a limited number of available measures, and most have been developed in the context of US education. The present study aimed to evaluate the psychometric properties of a test assessing the early development of numerical skills in the context of French preschool and kindergarten education. The test, which assessed key numerical abilities (i.e. encoding and decoding quantities, matching sets, comparing quantities, anticipating the result of an action), was based on a distinction between number used as a tool and number used as an object. Participants were children between the ages of 3 and 6. The results indicate that the test has good concurrent validity (i.e. correlates with an independent measure of math skills), good predictive validity (i.e. correlates between grades through the years of preschool and kindergarten), and that its structure can be explained by a Tool-Object dialectic framework. We argue that the test may be a useful tool for teachers to monitor early numeracy development, particularly in French-speaking classrooms.

ARTICLE HISTORY
Received 26 June 2024
Accepted 8 March 2026

KEYWORDS
Assessment tool; didactical analyses; preschool; tool-object dialectic; early numeracy

Introduction

The ability to understand and manipulate numbers is essential for both daily life and the workplace in our modern society. However, there are significant differences in numeracy skills between individuals (OECD 2023). An important finding from the past few decades of research is that those individual differences emerge early, even before children enter 1st grade (Dowker 2008). Furthermore, studies consistently find that early individual

CONTACT Marie-Line Gardes marie-line.gardes@hepl.ch Haute École Pédagogique du Canton de Vaud, UER MS, Avenue de Cour, 33, 1014 Lausanne, Switzerland

differences significantly predict later math achievement, from middle to high school (Duncan et al. 2007; Nguyen et al. 2016; Watts et al. 2014). A major explanation for this finding is that math skills are cumulative, in the sense that more advanced knowledge consistently builds on previous knowledge (Bailey et al. 2017; 2018). This naturally suggests that early numeracy knowledge set the stage for later math achievement (Torbeyns, Gilmore, and Verschaffel 2015). This also suggests that interventions that aim to reduce math difficulties may be particularly beneficial to children when they focus on foundational skills, during the period of early childhood education (Duncan and Magnuson 2013).

Previous tests of early numerical abilities developed in a US context

Early interventions require teachers and practitioners to be able to evaluate children's early numerical skills in a concise, reliable, and valid manner. Yet only a limited number of tests measuring early numerical skills (i.e. from preschool to kindergarten) have been empirically validated (Outhwaite et al. 2024). Most of these tests, such as the Woodcock–Johnson Tests of Achievement (Woodcock, McGrew, and Mather 2001), the Test of Early Mathematics Ability (Bliss 2006), the Preschool Numeracy Indicators (Floyd, Hojnoski, and Key 2006), the Tools for Early Assessment in Mathematics (Ginsburg and Baroody 2003; Sjoe et al. 2019), and more recently, the International Development and Early Learning Assessment (IDELA) – Emergent Numeracy (Save the Children 2019), provide relatively broad measures of math skills, mainly relying on scores that encompass a range of numerical skills. As noted by Purpura and Lonigan (2015), such broad measures may give relative estimates of children's skills relative to their peers, but do not make it easy to identify specific strengths and weaknesses in each child. Yet, this information is critical in a context of response to intervention, which aims to identify and support students who are struggling in specific areas by tailoring increasingly intensive interventions to those who do not respond to initial interventions (Fuchs and Fuchs 2006).

More recently, studies have attempted to develop tests of early numerical skills that may provide specific measures of different math skills in addition to a broad composite measure. For example, Li and colleagues (2020) recently developed a diagnostic cognitive test to specifically assess problem solving in children from 3 to 6. Similarly, the research-based early mathematics assessment (REMA) has been developed to measure a range of math competencies in children from 3 to 8 (Clements, Sarama, and Liu 2008; Dong et al. 2021; Weiland et al. 2012). Purpura and Lonigan (2015) have also developed the Preschool Early Numeracy Scales (PENS), which assesses a number of skills in children from 3 to 5 (see also (Purpura et al. 2015)). Both the REMA and the PENS have several strengths, including an extensive assessment of reliability and validity. Administration of each test can also relatively brief (a short version of the REMA has been developed (Weiland et al. 2012)).

However, as with many broader measures of math skills mentioned above, it is important to keep in mind that the REMA and the PENS have been developed in the context of US education. Therefore, it is unclear to what extent the scales are relevant and applicable in non-US contexts. Indeed, in contrast to the US where there is a wide heterogeneity in school (and especially preschool) systems (with formal education mainly starting in kindergarten), some countries have a more homogeneous formal standardised education system. In France, for instance, formal education is compulsory from the age of 3 and children attend a specific school called 'école maternelle' from age 3 to age 6 (before entering 1st grade in elementary school). In contrast to many US preschools, children attending the 'école maternelle' are instructed by trained teachers who follow a nation-wide math curriculum. Although domains of competence in the French national curriculum certainly overlap with the developmental progression underlying tests such as the REMA or the PENS, the fact that these tests were not developed in a French context make them more difficult to use in French classrooms (notably by teachers).

Previous French-language tests of early numerical abilities

A few batteries assessing early math skills have been validated in French children. These notably include the Test diagnostique des compétences de base en mathématiques (TEDI-MATH) (Grégoire, Noel, and Nieuwenhoven 2003) and the EXAMATH 5–8 (Lafay and Helloin 2016). However, these batteries, which have been largely designed to help diagnose learning disabilities in a clinical context, have two main shortcomings. First, they are relatively lengthy to administer, which does not make them appropriate to a classroom context. Second, they are both targeted to children who are older than 4 (5 for the EXAMATH 5-8) and therefore would not allow teachers and practitioners to assess children longitudinally from age 3 to age 6 (i.e. the age span of the 'école maternelle').

Recently, Thomas et al. (2021) developed a French-language early numeracy assessment that appears more appropriate for evaluating math skills in a classroom context. The test covers several dimensions of early numeracy, including number identification, number naming, verbal counting, one-by-one-counting, cardinality understanding, set comparison, subitising, set to numeral matching, and story problems. It can also be used to test children's math skills in both Kindergarten and pre-K. Yet, although the authors demonstrated that the test has good psychometric properties, it failed to identify separate math subskills (e.g. symbolic versus non-symbolic skills). The range of skills investigated is also limited, such that the test does not evaluate the ability to match two sets of equal cardinalities and includes few measures of problem-solving skill. Finally, the test has only been validated with children in Kindergarten and pre-K, which does not cover the span of the 'école maternelle' in France (which begins one year before pre-K at age 3).

In the present study, our goal is to follow-up from previous attempts (Courtier et al. 2021; Darnon and Fayol 2021; Thomas et al. 2021) and propose the Early NUmerical DIdactical Test (ENUDIT), a French-language test measuring a range of early numerical skills in children from the age of 3 to the age of 6 (i.e. from the beginning to the end of the 'école maternelle' period). We notably aimed to design a relatively brief test that can be used by teachers in the classroom to assess different aspects of number knowledge. We also aimed to evaluate the psychometric properties of this test. Critically, all of the tests detailed above are designed from the perspective of cognitive psychology. In line with recent attempts in France to develop collaborations between researchers in psychology and mathematics education, the ENUDIT builds on research in both psychology and mathematics education, so that it is likely to be accepted by both research communities. We detail below the framework upon which the ENUDIT is built.

The tool-object dialectic framework

The design of a test of early numerical abilities requires a detailed characterisation of the knowledge involved in learning the concept of number. This characterisation can come from different approaches: epistemological, cognitive, or didactic. It is also important for this characterisation to be multidimensional, i.e. to provide sufficient detail to accurately assess distinct numerical subskills through adequate item sampling for each component. An interesting conceptualisation of the way mathematical knowledge is constructed is the tool-object dialectic (Douady 1991, 2003), which comes from the field of mathematics education.

The tool-object dialectic model (Douady 1991; 2003) describes the teaching-learning relationship of a concept as well as the production and development of that conceptual knowledge from the student. This model relies on the idea that, as students learn a given concept (e.g. number), that concept cyclically alternates between having the status of a tool for solving a problem and having the status of an object in the construction of an organised form of knowledge. Take for example a concept such as factorisation. It can first be introduced as an essential tool for solving problems that would otherwise be unsolvable. However, it can then also become a topic of study in its own right, with dedicated chapters on advanced factorisation techniques aimed at developing specific expertise in this field. Finally, factorisation can regain its status as a tool to tackle more complex problems (in which case it will be one tool among others for solving problems).

The tool-object dialectic model has been used in a variety of contexts in mathematics education. For example, it has been applied to the domain of elementary algebra (Grugeon-Allys 2016) and has allowed for a detailed description of different types of algebraic tasks (Bosch and Gascón 2005). The model has also been used to probe the content and modalities of a particular teaching and how such content and modalities affect students' learning. For example, Modeste (2020) used the tool-object dialectic model to analyze the concept of algorithm in the context of mathematics teaching in France. Finally, the tool-object dialectic model has supported the design of specific pedagogical sequences (e.g. (Douady 1984; Douady and Perrin-Glorian 1989; Kouki, Belhaj Amor, and Hachaichi 2016)). Here we propose to use this tool-object dialectic model to characterise the multiple dimensions of early number abilities.

Indeed, number acquisition can be analyzed with the framework of the tool-object dialectic model. Specifically, at times number can be used by children as a tool to solve a problem. According to Douady (1991), a concept is used as a tool when 'the interest is focused on its use for solving problems. A tool is involved in a specific context, by someone, at a given time'. (Douady 1991, 115). Imagine, for example, that a child is presented with a set of garages and told that there are several cars at the other side of the classroom. The child is then asked to go get just enough cars so that each car has a garage. That child is clearly not required to use the concept of number to solve this problem. For instance, they can use one-to-one correspondence or make as many trips as there are garages. A critical observation here is whether the child will leverage the concept of number on their own to solve the problem (i.e. the problem can be solved by counting the set of garages). Although this situation is a classic example of using number as a tool, it is not unique in giving number the status of tool. For example, this is also the case when a child must compare quantities or anticipate what will be a resulting quantity after a transformation such as an addition or a subtraction. These three types of tasks (i.e. matching sets, comparing quantities, anticipating the result of an action) are those in which the concept of number may appear in the resolution process as one element among others (Margolinas and Wozniak 2012). For instance, when comparing quantities, children may consider using the concept of number for counting, but they must then compare the two quantities obtained.

At other times, however, number is clearly the specific object of study for children. For example, number clearly has the status of an object when a child quantifies a set of items. An observer (e.g. a teacher) may explicitly be interested in assessing how well the concept of number is acquired and what properties are known. This is also the case when children build sets of items given a specific cardinal. In other words, a concept is considered an object whenever an observer makes explicit that it needs to be used and retained as the sole element in the problem-solving process.

Early numerical skills and foundations for the ENUDIT

What type of knowledge constitutes early numerical skills? A number of studies coming from both numerical cognition and math education have now described the main components of early numerical skills (Baroody, Lai, and Mix 2006; Fayol and Seron 2005; Fuson 1988; Margolinas and Wozniak 2012, 2014; Purpura and Lonigan 2013; Sarama and Clements 2009). Broadly speaking, young children may use the concept of number in three different ways (Croset and Gardes 2019, 2020): as a measure of a quantity (3 tokens), as a measure of position (the 3rd position in an ordered list) and as a value (3) that is not practically connected to quantities and units of measurement (i.e. detached from concrete situations). Thus, developing an understanding of number requires an integration of these three perspectives: number as quantity, number as position and number decontextualised. The ENUDIT will focus on assessing number as quantity, given that this aspect is predominant in preschool and kindergarten. For instance, the French kindergarten curriculum conceptualises ten types of numerical tasks to be mastered by children: six focus on the concept of number as a quantity – determining the cardinality of a collection, making a collection of items given a cardinal number, building a collection with the same cardinality as a given collection, comparing the cardinalities of two sets of items, quantifying a collection after an addition or subtraction of one or more items, and composing and decomposing numbers. Only two numerical tasks in the curriculum focus on the concept of number as position – determining a position and comparing positions – while just one numerical task focuses on number in a context that is not practically connected to quantities and units of measurement (reading numbers written with digits). Note that the curriculum also divides these different types of numerical tasks into two categories: 'using numbers' and 'studying numbers', which corresponds to the tool-object dialectic (Douady 1986).

Based on a long-standing line of research in both math education and cognitive psychology (Baroody, Clements, and Sarama 2019; Clements, Baroody, and Sarama 2013; Gilmore et al. 2018; Margolinas and Wozniak 2012), Croset and Gardes (2019) distinguished five main types of tasks that young children need to work on in order to acquire the concept of number as a measure of quantity. First, they may encode a quantity. This involves determining the cardinal (i.e. the number of elements) of a collection.

Different procedures can be used to determine that cardinal, such as subitising or counting one-by-one. The cardinal can also be given by using either a verbal code or a written code. Second, children may decode a quantity. That is, they may construct a collection of items given a cardinal number. Decoding a quantity is an operation that is the reverse to encoding a quantity. Again, the cardinal can be communicated using a verbal code (e.g. give me five tokens) or written code (give me 5 tokens). Third, children may match a collection. In other words, they may build a collection that has the same cardinality as a given collection. In contrast to decoding a quantity, matching a collection does not involve a correspondence between a symbolic code and a given cardinal. Rather, the quantity to be used to build a collection is conveyed by another collection. Fourth, children may compare quantities. That is, they may extract and compare the cardinality of two sets of items. Finally, children may anticipate the result of an action on a given quantity, either by combining two separate quantities or transforming a given quantity by the addition or subtraction of another quantity (Riley, Greeno, and Heller 1983; Vergnaud 1982). Each of the five tasks described above provides the opportunity for number to be either an object of study or a tool for solving a problem, as described in the previous section. Specifically, two types of tasks require the explicit use of numbers in their resolution procedure (i.e. encoding and decoding quantities), while three other types of tasks integrate numbers as a possible, but not isolated, element in their resolution procedure (i.e. matching sets, comparing quantities, anticipating the result of an action) (Margolinas and Wozniak 2012).

The current study

Given the small number of assessment tools dedicated to assessing early number acquisition outside of a US context, the goal of the present study was to develop and validate a novel early numerical assessment test in a French context. We had two main criteria in mind while developing the test. First, the test had to assess number knowledge over the course of the French 'école maternelle', i.e. from the age of 3 to age of 5. Second, the test had to cover different dimensions of number comprehension, allowing for an independent assessment of subskills. In particular, we were interested in the theoretical distinction of number used as a tool and number used as an object. In the following, we describe the content of the ENUDIT (Early NUmerical DIdactical Test) and evaluate its psychometric properties by assessing its concurrent and predictive validity. We also evaluate to what extent the data support a partition of early number skills in terms of number used as tool and as object.

Materials and methods

Participants

Participants came from two diverse samples recruited in two French schools of the Lyon area in France. The first school was a public school with nine classrooms. The school was located in a disadvantaged neighbourhood of Lyon, as shown by a median equivalized disposable income of €15,208, which is largely below the national median equivalized disposable income (i.e. €20,809, https://www.insee.fr/fr/statistiques/4195239). As described in Courtier et al. (2021), about 33% of individuals living in this area have an income that is below the French poverty threshold. The school belonged to the French 'Reinforced Priority Education Network', a network of schools in which students and teachers benefit from various types of support aimed at reducing inequalities and enhancing academic achievement of underprivileged students. The second school was a private school with two classrooms. In contrast to the first school, that second school was located in a more advantaged area, as demonstrated by a median equivalized disposable income of €30,020. Only 8% of individuals living in this area have an income that is below the poverty threshold. Background information was also directly collected from parents of children, though we were only able to gather information from a subset of families (see Courtier et al. 2021, for details). In the first (public) school, monthly household income of parents ranged from less than €1,500 to between €5,000 and €7,500, with an average between €1,500 and €3,000. In the second (private) school, household income of parents ranged from between €1,500 and €3,000 to more than €7,500 a month, with an average between €3,000 and €5,000. Therefore, across both samples, the socio-economic status (SES) of children was diverse. Note that data on participant ethnicity were not collected because the collection of such data is in principle illegal in France and would require an exceptional waiver from government agencies that we did not seek. The data set was collected between October 2015 and June 2019 as part of another research project (Courtier et al. 2021). The experiment was approved by the local school board and was performed in accordance with the ethical standards established by the Declaration of Helsinki.

Two samples were considered in the current study. The first sample (i.e. Sample #1) was composed of children (coming from both the public and private school) from three different grades: Petite Section (PS, corresponding to two years before kindergarten), Moyenne Section (MS, corresponding to one year before kindergarten), and Grande Section (GS, corresponding to kindergarten). PS included 225 children who were evaluated at the beginning of their first year of preschool (average age: 3.31 years, 111 girls). MS included 247 children who were evaluated in the middle of their second year of preschool (average age: 4.61 years, 114 girls). GS included 172 children who were evaluated at the end of their third year of preschool (average age: 5.96 years, 86 girls). Sample #1 was used for the confirmatory factor analysis.

The second sample (i.e. Sample #2) was a subset of Sample 1. The children of this second sample are those whom we have been able to follow longitudinally during the three years of preschool and kindergarten. Sample 2 was composed of 75 students (44 girls). These children were tested three times: at the beginning of PS, in the middle of MS, and finally at the end of GS. This sample was notably used to measure predictive validity.

Measures

To evaluate concurrent validity, all students were evaluated with the widely used Applied Problems subtest of the Woodcock-Johnson III (WJ-III) (Woodcock, McGrew, and Mather 2001) in addition to the ENUDIT. The Applied Problems subtest form the WJ-III, which was translated into French, measures quantitative reasoning as well as mathematical knowledge and competencies (Woodcock, McGrew, and Mather 2001).

Early items notably call upon basic numerical concepts (e.g. counting, performing simple addition and subtraction). The test is untimed and composed of 63 items of increasing difficulty. After six consecutive errors, the test is stopped. This test was chosen because it is widely considered an established and valid test of overall math competence, as shown by its wide use in studies of early math knowledge (Blair and McKinnon 2016; Espy et al. 2004; Gormley, Phillips, and Gayer 2008; Lillard 2012; Lillard et al. 2017; Lillard and Else-Quest 2006).

The ENUDIT was based on the theoretical framework of early numerical development described in the introduction. In line with the framework proposed by Croset and Gardes (2019), the tests included five types of tasks, from T1 to T5, corresponding to tasks listed in Table 1. As is clear from the table, two types of tasks allow number to be used as an object and three types of tasks require number to be a tool (Douady 1986). The tasks were developed on the basis of skills tested in other existing international assessment tools, including PENS (Purpura and Lonigan 2015) and another recent French-language early numeracy assessment (Thomas et al. 2021). The test was also in line with the French kindergarten curriculum (MEN 2015), which indicates the skills that children should acquire each year of kindergarten. Each task was designed so that it was relatively easy to administer, using simple and straightforward instructions. The vocabulary used in the tasks was kept simple, as the words were among the most frequent[1] in the French language (e.g. add, hand) and in the usual vocabulary[2] of children aged 3–6, according to the French ministry of Education (e.g. token, hand, add). Each task used neutral materials (e.g. tokens) that children could manipulate easily (Carbonneau et al. 2013). Table 1 lists the 5 types of tasks with their instructions. For each type of tasks, we used items of increasing difficulty. For example, during T1, children started by

Table 1. Five types of tasks included in the ENUDIT.

Type of task	Task	Number of items	Values of items	Number used as
Encoding	T1 – Determine the cardinal of a set.	3 items	3, 7, 11	Object
Decoding	T2 – Construct a set with a given cardinality	2 items	5, 12	Object
Matching	T3 – Match a set of the same number of elements as another set (i.e. 'In one trip, get just enough red tokens to put on top of the blue tokens. You have to take that many and you can't go back for more')	2 items	5, 12	Tool
Comparing	T4 – Compare the cardinality of two sets (i.e. 'In which pile are there the most tokens?')	2 items	2 and 3, 8 and 9	Tool
Anticipating (combining)	T5a – Solve problems that involve combining sets (i.e. 'There are x tokens in my left hand. There are y tokens in my right hand. How many do I have in total?' or 'There are x tokens in my left hand. There are some tokens in my right hand. There are y tokens in total. How many tokens are in my right hand?')	2 items	$3 + 4 = ?$ $2 + ? = 6$	Tool
Anticipating (transforming)	T5b – Solve problems that involve transformation (i.e. 'There are x tokens in my hand. If I add y tokens. How many tokens will there be?' or 'There are x tokens in my hand. If I take away y tokens. How many tokens will there be?')	8 items	$3 + 1 = ?$; $3 + 2 = ?$; $3 - 1 = ?$; $3 - 2 = ?$; $6 + 1 = ?$; $6 + 2 = ?$; $6 - 1 = ?$; $6 - 2 = ?$	Tool

identifying the cardinal of a set of 3 tokens before identifying the cardinal of 7 tokens and eventually 11 tokens. The number of items successfully completed by children on that specific task therefore ranged from 0 to 3 (reflecting their success in identifying the cardinals). For each type of tasks, a percent completion was calculated, such that the minimum score was always 0 and the maximum score was always 1. In total, children were tested on 19 items divided into 5 types of tasks. Therefore, the total score was between 0 and 6. The cardinal values children were presented with were chosen based on the French school curriculum and an analysis of school textbooks (i.e. around 3 for PS, around 6 for MS and around 10 for GS).

Testing procedure

Children were tested individually in a quiet room of their preschool. The tests were administered by different experimenters (graduate students, research assistants and undergraduate students) as part of a broader evaluation of cognitive and academic skills that consisted in five sessions of approximately 15–20 min. Administration of the Applied Problems subtest took approximately 10 min. Administration of the ENUDIT took approximately 20 min. Order of testing sessions was randomised on a child-to-child basis. No feedback was given to children during testing. Teachers and parents were blind to the tests until the end of data collection.

Analytic procedures

After reporting descriptive statistics for the five ENUDIT tasks (see Table 2), we evaluated the concurrent and predictive validity of the ENUDIT, as well as the extent to which its structure can be described by the tool-object distinction described in the Introduction. First, concurrent validity was evaluated on Sample #1 by comparing the overall ENUDIT score measured at a given grade (PS, MS, and GS) with the score obtained on the WJ-III Applied Problems subtest at the same grade, using independent Pearson correlations (with the cor.test function in R). Because longitudinal data were available on a subset of participants (i.e. Sample #2), we also evaluated the relation between ENUDIT total scores and WJ-III Applied Problems scores within that subsample. To take into account the dependency in the data (i.e. each child was tested three times), this was done using a linear mixed model with the ENUDIT total score as dependent variable, adjusting for WJ-III Applied-Problems scores, grade level (PS, MS, and GS) and including an individual random effect for each child. We used the lme function from the nlme

Table 2. Descriptive statistics for the five types of tasks of the ENUDIT.

	PS ($n = 225$)			MS ($n = 247$)			GS ($n = 172$)		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
T1. Encoding	0.087	0.188	[0; 1]	0.517	0.347	[0; 1]	0.816	0.256	[0.3; 1]
T2. Decoding	0.040	0.151	[0; 1]	0.423	0.393	[0; 1]	0.881	0.251	[0, 1]
T3. Matching	0.020	0.119	[0; 1]	0.119	0.261	[0; 1]	0.378	0.3	[0, 1]
T4. Comparing	0.467	0.436	[0; 1]	0.907	0.220	[0; 1]	0.968	0.123	[0.5, 1]
T5a. Anticipating (combining)	0.151	0.235	[0; 1]	0.310	0.335	[0; 1]	0.619	0.349	[0, 1]
T5b. Anticipating (transforming)	0.092	0.129	[0; 0.75]	0.342	0.314	[0; 1]	0.749	0.269	[0,1]
Total score	0.857	0.662	[0; 3.21]	2.618	1.184	[0; 5.75]	4.411	0.956	[1.33; 6]

Notes: SD, Standard deviation. Scores for each task range from 0 to 1. Total score ranges from 0 to 6.

package in R. The linear correlation adjusted on Class and Child was tested with a Wald test.

Second, predictive validity of the ENUDIT was measured within the longitudinal sample (i.e. Sample #2). Specifically, we assessed the strength of the relation between (1) total ENUDIT scores in PS and total ENUDIT scores in GS, as well as (2) total ENUDIT scores in PS and WJ-III Applied Problems scores in GS. A linear regression model was applied with GS score as dependent variable and PS score as independent variable. The lm function in R was used to estimate the parameters of the model. Predictive validity was tested with a Wald test of the regression coefficient (slope of the regression). Each model was validated through plots of residuals.

Finally, we tested the extent to which the tool-object distinction may explain the structure of the test with the full sample of participants (i.e. Sample #1). A confirmatory factor analysis (CFA) was performed with structural equation models, using the function cfa of the lavaan R package. Performance on the 19 items was entered in the CFA. Comparative Fit Indices (CFI) values greater than 0.95 were considered as indicating a good fit and Standardised Root Mean Square Residuals (RMSEA) value less than or equal to 0.05 were also considered as a good fit. In a first CFA, we tested a one-factor model with no relation between the tasks. This model was considered the baseline. A two-factor model was then applied. This two-factor model contained two variables (i.e. Tool and Object), each containing several subitems. The variable *Object* had two subitems corresponding to the first two tasks, i.e. Encode (T1) and Decode (T2). The variable *Tool* had three subitems corresponding to the other tasks, i.e. Match (T3), Compare (T4), and Anticipate (T5 and T6). These models were applied at each grade level (PS, MS, and GS).

Results

Descriptive statistics, item analyses and between task correlations

Descriptive statistics for each type of tasks and each grade level are shown in Table 3. As can be seen from the table, performance increased from PS to GS, though there was considerable variation in scores for each task. Scores associated with coding and decoding quantities steadily increased across the 3 levels. Matching quantities remained difficult through the years, even for children at the GS level. On the contrary, comparing quantities was associated with relatively high accuracy in MS and GS. Children clearly progressed with tasks requiring anticipation of quantities from PS to GS, in fairly similar ways between the combine and transform versions of the task. Table 3 shows the strength of the bivariate correlations between each type of tasks at each grade level. Correlations range from small to medium. Notably, tasks T5a and T5b correlated at all grade levels.

Concurrent validity of the ENUDIT

Prior to examining concurrent validity, we evaluated the internal consistency of the ENUDIT tasks. Cronbach's alpha coefficients ranged from marginal to good across tasks ($\alpha = 0.64$–0.87). The Anticipating tasks demonstrated excellent internal consistency ($\alpha = 0.87$), while Encoding, Decoding, and Comparing tasks showed acceptable consistency ($\alpha = 0.73$–0.80). The Matching task exhibited marginal internal consistency ($\alpha = 0.64$), which warrants cautious interpretation of subsequent analyses involving this construct.

Table 3. Pearson correlations between the ENUDIT tasks by grade.

	PS ($n = 225$)					MS ($n = 247$)					GS ($n = 172$)				
	T1	T2	T3	T4	T5a	T1	T2	T3	T4	T5a	T1	T2	T3	T4	T5a
T1															
T2	**0.29**					**0.54**					**0.26**				
T3	0.09	−0.04				**0.24**	**0.34**				0.10	**0.16**			
T4	0.10	0.02	0.01			0.11	**0.16**	0.09			**0.15**	0.11	0.04		
T5a	0.09	0.08	−0.03	0.07		**0.22**	**0.24**	0.12	0.06		**0.24**	**0.15**	0.13	0.12	
T5b	0.10	0.11	**0.15**	0.13	**0.15**	**0.42**	**0.50**	**0.30**	**0.19**	**0.24**	**0.22**	**0.27**	**0.31**	**0.18**	**0.34**

Notes. Values in bold are significant ($p < 0.05$). T1, Encoding, T2, Decoding, T3, Matching, T4, Comparing, T5a, Anticipating (combining), T5b, Anticipating (transforming).

We then evaluated the concurrent validity of the ENUDIT by measuring to what extent its total score correlated with the WJ-III Applied Problems subtest when assessed at the same time (see Figure 1). Pearson correlations were significant at the three grade

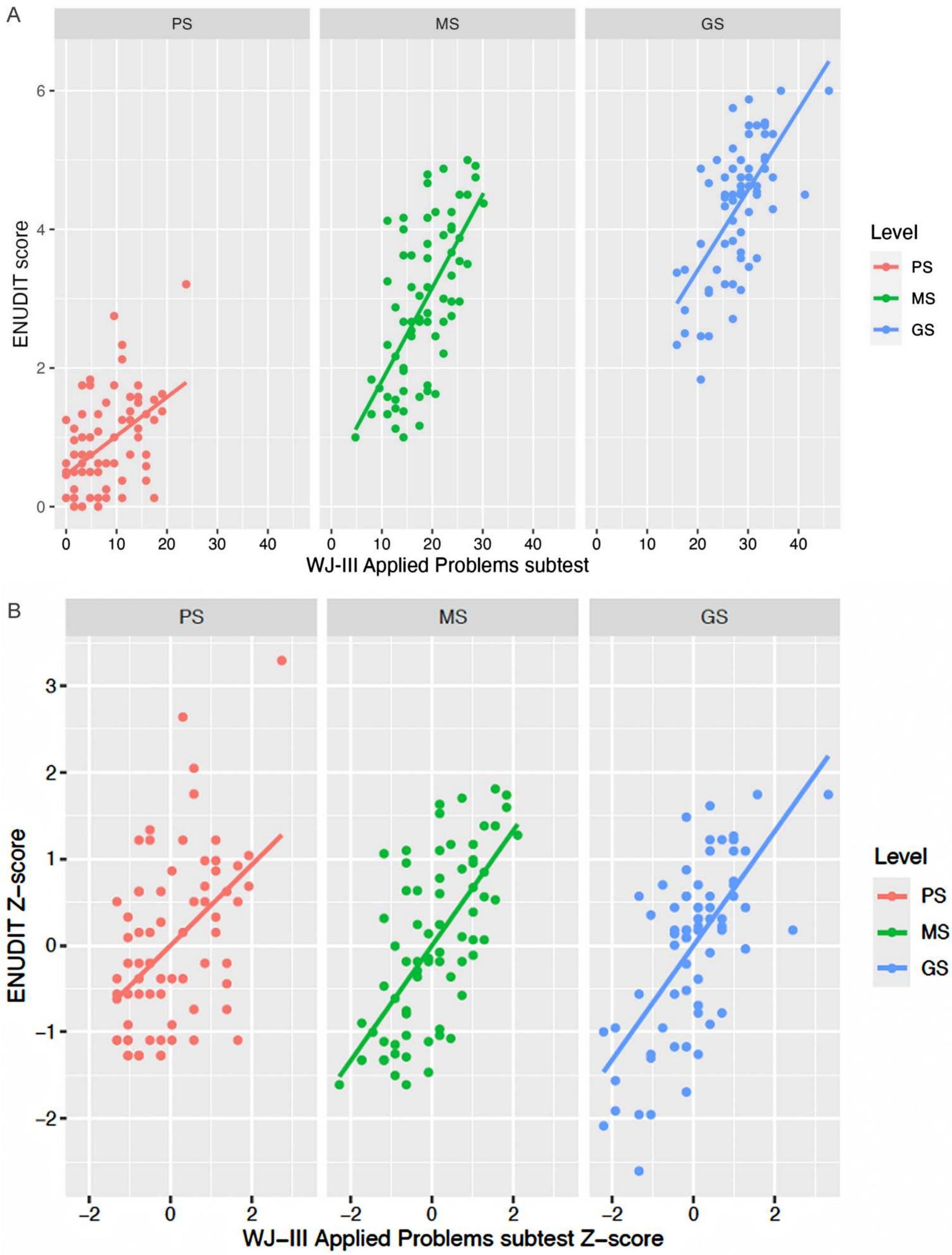

Figure 1. Concurrent validity. A-Correlation between the total score of the ENUDIT and the score of the WJ-III Applied Problems subtest at each grade level across the whole sample (i.e. Sample #1). B-Correlation between the Z-score of the ENUDIT and the Z-score of the WJ-III Applied Problems subtest at each grade level across the whole sample (i.e. Sample #1).

levels and ranged from medium to large, with the smallest correlation observed at the PS level ($r = 0.467$, $p < 0.001$) and the largest at the MS level ($r = 0.666$, $p < 0.001$). The correlation for the GS level was also large ($r = 0.661$, $p < 0.001$).

We also evaluated the strength of the correlation between the ENUDIT total score and the WJ-III Applied Problems score within the longitudinal sample (i.e. Sample #2). This was done with a linear mixed model, considering the WJ-III as dependent variable and the ENUDIT total score as the predictor, while adjusting for the level (PS, MS, and GS) and including an individual random effect for each child. The analysis showed a significant effect of the ENUDIT total score on the WJ-III score in PS ($t(139) = 3.11$, $p = 0.002$), MS ($t(139) = 13.80$, $p < 0.001$), and GS ($t(139) = 24.73$, $p < 0.001$).

Predictive validity of the ENUDIT

The predictive validity of the ENUDIT was assessed in the subsample of participants for which we had longitudinal data from PS to GS (Sample #2). A first linear model showed a significant effect of ENUDIT total score in PS on ENUDIT score in GS ($t(69) = 2.747$, $p = 0.007$, $R2 = 0.085$), indicating that children who were the most accurate on the ENUDIT in PS were also the most accurate in GS (see Figure 2). A second linear model showed a significant effect of ENUDIT total score in PS on WJ-III Applied problems score in GS ($t(69) = 2.612$, $p = 0.011$, $R2 = 0.077$), indicating that children's ENUDIT scores in PS were predictive of later overall math competence (as measured by the WJ-III) in GS (see Figure 3). Therefore, the test appears to have good predictive validity through the preschool and kindergarten years.

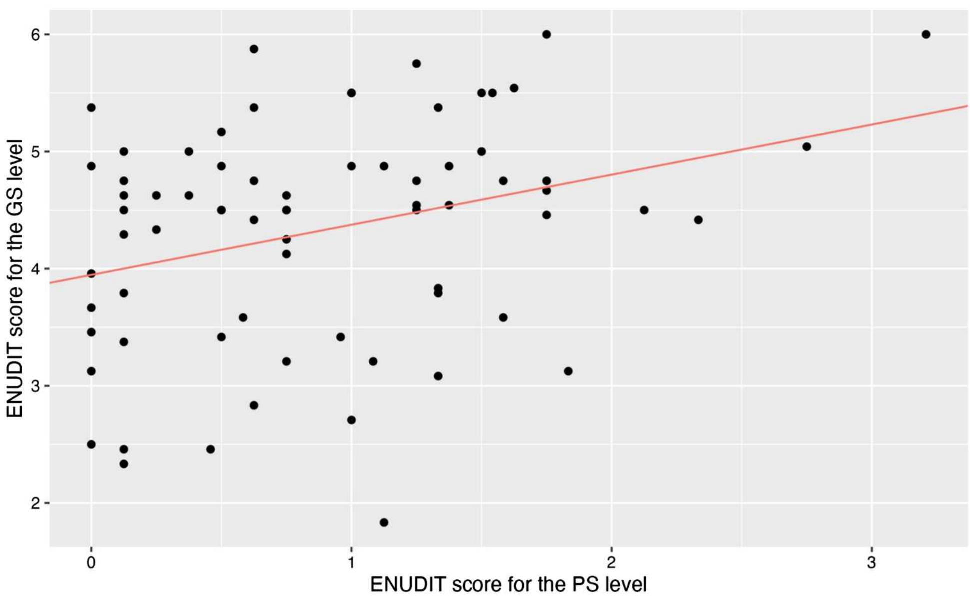

Figure 2. Predictive validity. Correlation between the total score of the ENUDIT in PS and the total score of the ENUDIT in GS within the longitudinal sample (i.e. Sample #2).

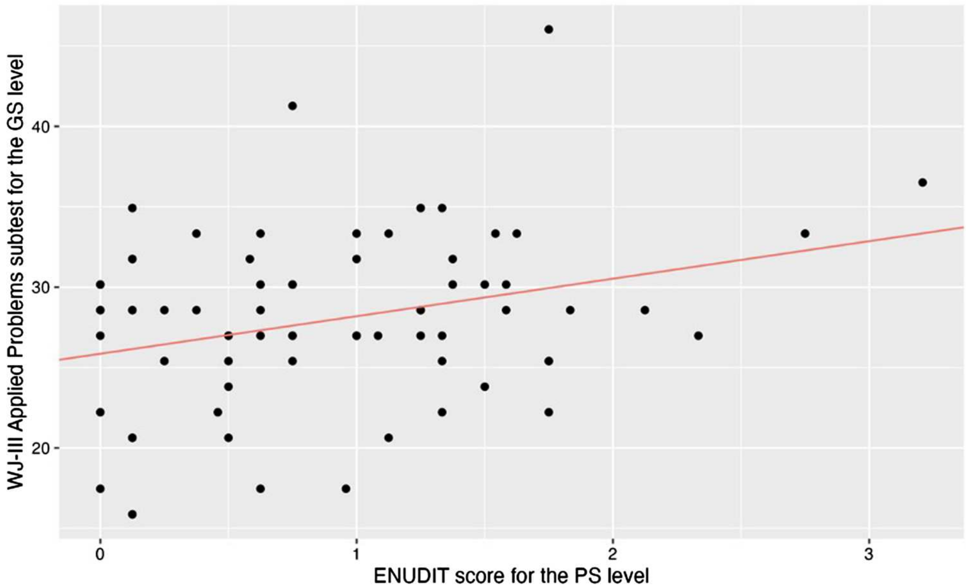

Figure 3. Predictive validity. Correlation between the total score of the ENUDIT in PS and the WJ-III Applied problems score in GS within the longitudinal sample (i.e. Sample #2).

Does the ENUDIT support a tool-object dissociation?

Finally, we performed a series of confirmatory factor analyses (CFAs) on the ENUDIT tasks to evaluate the extent to which the Tool-Object Dialectic framework may explain the structure of the test. Baseline one-factor models indicated a poor fit between the data and the model in PS (CFI = 0.278; RMSEA = 0.117), MS (CFI = 0.780; RMSEA = 0.088), and GS (CFI = 0.717; RMSEA = 0.082). However, two-factor models (with the factors Tool and Object) indicated a fit between the data and the model that was acceptable in PS (CFI = 0.811; RMSEA = 0.065) (see Figure 4(a)) and MS (CFI = 0.926; RMSEA = 0.051) (see Figure 4(b)), and good in GS (CFI = 0.944; RMSEA = 0.038) (see Figure 4(c)). Therefore, there was support for the Tool-Object Dialectic framework as a model explaining the structure of the ENUDIT.

Discussion

In the present study, we aimed to develop the ENUDIT, a progress monitoring test assessing early numerical abilities in French preschoolers. The test measures several key early numerical competences and is based on a framework coming from math education, the Tool-Object Dialectic framework (Douady 1991). The ENUDIT was designed to test children from the age of 3 to the age of 6. First, we found that the ENUDIT total score correlated with the score on the WJ-III Applied Problems, a broad measure of math skills that has been in previous research (Woodcock, McGrew, and Mather 2001). Second, the ENUDIT total score measured at age 3 correlated with the total score obtained at the end of kindergarten (approximately age 6). It also correlated with the score on the

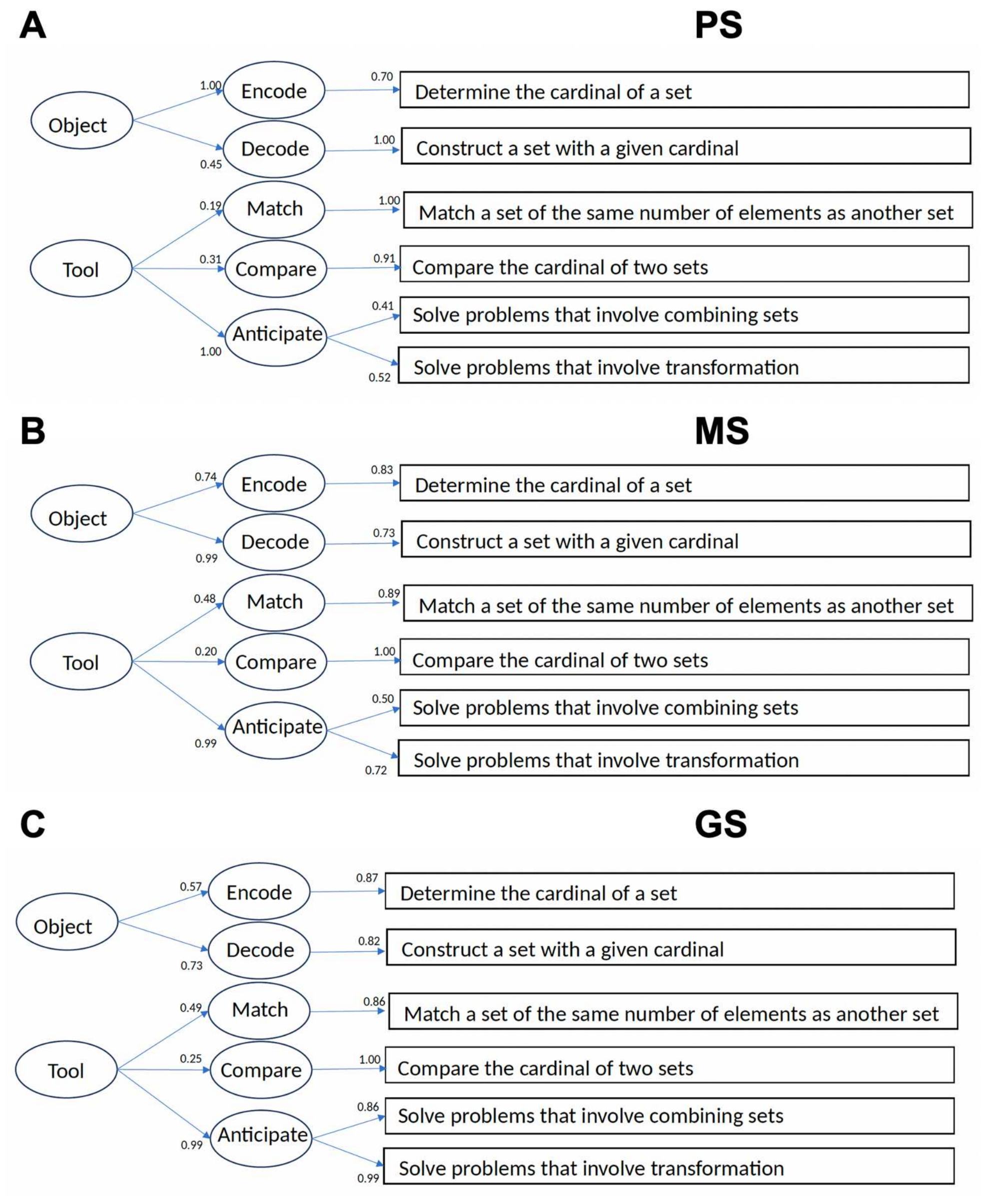

Figure 4. Confirmatory factor analyses (two-factor models) of the ENUDIT for PS (A), MS (B), and GS (C). Ovals indicate latent variables; rectangles indicate measured variables. All values are standardised coefficients.

WJ-III Applied Problems measured at the end of kindergarten. Together, these results indicate that the ENUDIT is a valid measure of numerical skills. Third, CFAs revealed that the structure of the data was better accounted by a two-factor model separating tasks in which number is considered a tool from tasks in which number is considered an object than for a one-factor model. This supports the idea that subtests of the ENUDIT may be sorted in these two categories (number as tool and number as object).

A test assessing key numerical skills

The ENUDIT allows for independent measures of relatively fine-grained numerical competencies. These include encoding and decoding quantities, comparing quantities, matching sets, and anticipating the results of actions on quantities. Psychological research has long recognised these skills as fundamental to the development of early numerical skills (Baroody, Clements, and Sarama 2019). Therefore, they are also central to other tests of early numerical abilities. For example, two US-based tests (Purpura and Lonigan 2015; Weiland et al. 2012) and a French test (Thomas et al. 2021) incorporate five of the six ENUDIT types of tasks: encoding, decoding, comparison, and anticipation tasks. Note that the matching task (building a collection that has the same cardinality as a given collection) is a relatively unique feature of the ENUDIT. We believe that this task is informative for assessing early numerical skills for two reasons.

First, the matching task requires one-to-one correspondence as a guarantee of numerical equivalence (i.e. Hume's principle). Carey and Barner (2019) highlight the critical role of Hume's principle in the conceptualisation of number, especially for understanding the cardinal principle. However, tests often assess one-to-one correspondence for its own sake (e.g. in a counting task, where the success criterion is the association of the number word with exactly one object) and do not evaluate it as a guarantee of set equality, which the matching task enables.

Second, the matching task assesses the spontaneous use of number. To succeed in that task, children must be able to mobilise numbers independently to first quantify the reference collection and then retrieve the same quantity of objects from another collection. However, young children do not always spontaneously use numbers in situations that require it (Brousseau 2012; Margolinas and Wozniak 2012; Wynn 1992), which is again a reason to use the task.

A test based on the tool-object dialectic framework

A critical aspect of the ENUDIT is that it has been conceived as a tool allowing for the categorisation of different numerical tasks. Thomas and colleagues attempted to categorise items based on formal and informal knowledge; however, their results do not support such a model. The ENUDIT draws from the Tool-Object dialectical framework (Douady 1986). Originating from the field of mathematics education, the model posits that learning and teaching mathematical concepts hinge on the interplay between these two dimensions. To our knowledge, ENUDIT is unparalleled among early numeracy tests in this conceptualisation of skills. Critically, the CFA results suggest that this framework helps with the categorisation of early numerical skills, facilitating independent assessments of the two dimensions and offering additional granularity. In essence, the ENUDIT enables educators to identify the strengths and weaknesses of children when numbers serve as tools or objects.

Several characteristics of the test make this tool suitable for use in educational settings. First, the test is short (20 min), allowing teachers to administer it without significantly disrupting teaching time. Second, the test is based on a common school culture, both in the choice of tasks (related to the curriculum) and in the formulation of instructions, which facilitates its adoption and use. Indeed, the questions and tasks proposed are aligned with current French teaching expectations and practices, making the results interpretable for the teacher. Third, its psychometric qualities, particularly its predictive validity, may enable teachers to reliably identify students who are likely to encounter future difficulties in the domain of number. Fourth, one of the key strengths of the test lies in its ability to differentiate difficulties of students based on the conception of number: as an object (e.g. as a collection of elements to be produced) or as a tool (e.g. used in matching activities). This distinction provides teachers with valuable information to guide their pedagogical actions more effectively. In fact, to assess whether a student is in difficulty, teachers can also compare their performance to the class average scores or to the reference averages available in Table 2. This comparison provides an objective basis for identifying students who need additional attention. Finally, the test may allow teachers to identify specific difficulties based on an a priori analysis of typical errors (see for example such an analysis in French here: https://osf.io/tvd2g). Such a remediation can consider the specific nature of the errors made by the student (Margolinas and Wozniak 2012). For instance, should children fail both the coding and matching tasks, remediation could target the coding task. Conversely, if the coding task is successful but the matching task is not (which, in our sample includes half of the GS students and over two-thirds of the MS students) then remediation could concentrate on that type of task. Remediation could be structured in steps, such as permitting children to undertake as many attempts as necessary, thereby helping them with the acquisition of Hume's principle (two sets are equal in number if and only if they can be placed in perfect one-to-one correspondence). In sum, the ENUDIT may offer teachers a quick, predictive, and targeted assessment of students' skills, allowing for early and adapted intervention based on a detailed understanding of the difficulties encountered.

A test adapted to the assessment of early numerical skills in the French classroom

To our knowledge, the majority of tests assessing early numeracy skills have been developed within the context of US preschool education (Outhwaite et al. 2024; Purpura and Lonigan 2015; Weiland et al. 2012). In contrast, the ENUDIT has its origins in French preschool education, making it particularly apt for evaluating early numerical skills in French classrooms. First, the test caters to children aged 3–6, aligning with the age range of France's 'école maternelle'. Second, the test is relatively brief (approximately 20 min) and can be administered by teachers and practitioners in the classroom. Third, it evaluates the key mathematical milestones delineated in the French preschool and kindergarten curricula. As previously noted, the French kindergarten curriculum organises tasks into ten types: six addressing number as quantity, two as position, and one in a way that is decontextualised. The ENUDIT tests all six tasks related to number as quantity, thus covering the comprehensive expectations regarding this aspect of number understanding, which is a central mathematical concept in French kindergarten.

However, the test could also be used in environments other than the French classroom, especially among French-speaking students when their curriculum parallels the French curriculum. For instance, Gardes et al. (2021) examined the curriculum for French-speaking Switzerland, noting that it focused solely on numbers as quantity and relies on the tool dimension within the Tool-Object framework (Douady 1986): 'In the first part of the cycle, numbers are not the focus of study in themselves, but serve as tools for naming, reading, and writing quantities in practical activities or learning contexts' (CIIP 2010). Therefore, the test may help in the evaluation of all types of tasks in the curriculum, such as counting a small collection of objects (Task 1 – Coding), comparing two collections (Task 4 – Comparing), creating a collection with a specified number of objects (Task 2 – Decoding), as well as modifying the quantity of objects in a collection and solving additive and subtractive problems through enactment, illustration, or manipulation (Task 5 – Anticipating). Although matching a set with the same number of elements as another set is not explicit in the curriculum, it is present in several tasks within the textbook available to educators. Therefore, this concept is taught and would probably be beneficial for educators to evaluate using the test, even if not explicitly included in the curriculum.

As a final note, administering a math test to young children can pose significant comprehension challenges due to individual differences in language abilities. The vocabulary used in the instructions is therefore important as it can potentially lead to bias in the assessment of numeracy skills (Purpura and Reid 2016). When designing the ENUDIT, we chose to rely on vocabulary that was simple and easily understandable for young children (e.g. hand, token, add, etc.). As such, we believe it makes it feasible to adapt the ENUDIT for use in other French-speaking countries with minimal changes to the vocabulary. On the other hand, several studies have highlighted the influence of the lexical properties of number words on number learning (e.g. Browning and Beauford 2012). However, these effects are largely related to learning the decimal system and are therefore mainly apparent for numbers greater than 10. Given that the ENUDIT test only covers numbers between 0 and 10 and that, to our knowledge, all number word systems have specific words to name these first numbers, we believe that the test can be used in its current form in different countries.

Limitations and future research

While we recognise that the ENUDIT addresses a gap in the literature by offering a novel, validated French-language assessment of early numerical skills, our study is not without limitations. First, the sample is limited to French children. Considering the potential application of the test in countries with similar curricula, gathering data from these countries would be valuable to evaluate the ENUDIT's structure and to determine if the Tool-Object distinction is fundamentally developmental or contextual.

Second, the ENUDIT places emphasis on the cardinal aspect of number, aiming for a comprehensive set of tasks in this area. As Croset and Gardes (2019, 2020) have articulated, a thorough comprehension of number skills also involves evaluating two additional aspects: number as a measure of position and number abstracted from tangible situations (i.e. number decontextualised). Notably, several studies suggest that processing the ordinal aspect of numbers is essential for the development of advanced numerical and symbolic abilities (Brannon and Van de Walle 2001; Bruce and Threlfall 2004; Lyons et al. 2014; Margolinas and Wozniak 2014). Yet, the PENS, the REMA, and the Thomas test do not encompass this dimension of number. Given that the ENUDIT is inspired by Croset & Gardes' work – which does include these aspects – future iterations of the ENUDIT could potentially integrate this component. It should be noted that the test also does not include spatial tasks assessing early geometric abilities.

Third, the ENUDIT was developed to measure milestones in early numerical skills as specified by the French preschool curriculum. Therefore, it currently evaluates only skills pertaining to basic number concepts. Extending the scope to include more complex skills, such as comprehension of the decimal system and place-value (the relation between the position of a digit and the number of groupings), could make for a more discriminating test.

Conclusion

We have developed the ENUDIT, a progress-monitoring tool designed to assess early numerical abilities in French preschoolers. The findings confirm that the ENUDIT is both a valid and reliable tool for evaluating numerical skills. Furthermore, these findings show the relevance of the Tool-Object model in the categorisation of early numerical skills, highlighting the practical implications of our research for education. The instrument is tailored to be user-friendly for kindergarten teachers and is suitable for children aged 3–6. It enables educators to consistently evaluate their students' knowledge at both the beginning and the end of each academic year in preschool and kindergarten settings. Such consistent assessment provides a benchmark for pinpointing students' challenges and subsequently customising instruction, particularly by implementing remedial strategies that are responsive to individual learning needs.

Notes

1. https://eduscol.education.fr/document/15655/download.
2. http://ien-neuilly.circo.ac-creteil.fr/IMG/pdf/2500_mots_themes.pdf.

Acknowledgements

This research was supported by funding from the Institut Carnot de l'Éducation under Grant [PR06]. We thank all students for their assistance in collecting the data. We are also grateful to all of the children who participated in this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The data that support the findings of this study are openly available in OSF at https://osf.io/4sze2/, reference ENUDIT.

ORCID

Marie-Line Gardes http://orcid.org/0000-0003-2448-2238
Marie-Caroline Croset http://orcid.org/0000-0003-2399-5883
Jérôme Prado http://orcid.org/0000-0003-3548-5610

References

Bailey, D. H., G. J. Duncan, C. L. Odgers, and W. Yu. 2017. "Persistence and Fadeout in the Impacts of Child and Adolescent Interventions." *Journal of Research on Educational Effectiveness* 10 (1): 7–39. https://doi.org/10.1080/19345747.2016.1232459.

Bailey, D. H., G. J. Duncan, T. Watts, D. H. Clements, and J. Sarama. 2018. "Risky Business : Correlation and Causation in Longitudinal Studies of Skill Development." *American Psychologist* 73 (1): 81–94. https://doi.org/10.1037/amp0000146.

Baroody, A. J., D. H. Clements, and J. Sarama. 2019. "Teaching and Learning Mathematics in Early Childhood Programs." In *Handbook of Early Childhood Care and Education*, edited by C. Brown, M. B. McMullen, and N. File, 1st ed., 329–353. Hoboken, NJ: Wiley Blackwell Publishing.

Baroody, A. J., M. Lai, and K. S. Mix. 2006. "The Development of Young Children's Early Number and Operation Sense and Its Implications for Early Childhood Education." In *Handbook of Research on the Education of Young Children*, edited by B. Spodek and O. N. Saracho, 2nd ed., 205–240. Lawrence Erlbaum Associates.

Blair, C., R. D. McKinnon, F. L. P. Investigators and Others. 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.

Bliss, S. 2006. "Test of Early Mathematics Ability–Third Edition." *Journal of Psychoeducational Assessment* 24 (1): 85–98. https://doi.org/10.1177/0734282905282839.

Bosch, M., and J. Gascón. 2005. "La praxéologie comme unité d'analyse des processus didactiques." In *Balises pour la didactique des mathématiques*, edited by A. Mercier and C. Margolinas, 107–122. Grenoble: La Pensée Sauvage.

Brannon, E. M., and G. A. Van de Walle. 2001. "The Development of Ordinal Numerical Competence in Young Children." *Cognitive Psychology* 43 (1): 53–81. https://doi.org/10.1006/cogp.2001.0756.

Brousseau, G. 2012. "Des dispositifs Piagétiens … aux situations didactiques." *Éducation et didactique* 6 (2): 103–129. https://doi.org/10.4000/educationdidactique.1475.

Browning, S. T., and J. E. Beauford. 2012. "Language and Number Values: The Influence of Number Names on Children's Understanding of Place Values." *Investigations in Mathematics Learning* 4 (2): 1–24. https://doi.org/10.1080/24727466.2012.11790310.

Bruce, R., and J. Threlfall. 2004. "One, Two, Three and Counting." *Educational Studies in Mathematics* 55 (1): 3–26. https://doi.org/10.1023/B:EDUC.0000017676.79430.dc.

Carbonneau, K. J., S. C. Marley, and J. P. Selig. 2013. "A Meta-Analysis of the Efficacy of Teaching Mathematics with Concrete Manipulatives." *Journal of Educational Psychology* 105 (2): 380–400. https://doi.org/10.1037/a0031084.

Carey, S., and D. Barner. 2019. "Ontogenetic Origins of Human Integer Representations." *Trends in Cognitive Sciences* 23 (10): 823–835. https://doi.org/10.1016/j.tics.2019.07.004.

CIIP. 2010. [Plan d'études romand]. www.plandetudes.ch.

Clements, D. H., A. J. Baroody, and J. Sarama. 2013. *Background Research on Early Mathematics. National Governor's Association, Center Project on Early Mathematics.*

Clements, D. H., J. H. Sarama, and X. H. Liu. 2008. "Development of a Measure of Early Mathematics Achievement Using the Rasch Model : The Research-based Early Maths Assessment." *Educational Psychology* 28 (4): 457–482. https://doi.org/10.1080/01443410701777272.

Courtier, P., M.-L. Gardes, J.-B. Van der Henst, I. A. Noveck, M.-C. Croset, J. Epinat-Duclos, J. Léone, and J. Prado. 2021. "Effects of Montessori Education on the Academic, Cognitive, and Social Development of Disadvantaged Preschoolers : A Randomized Controlled Study in the French Public-School System." *Child Development* 92 (5): 2069–2088. https://doi.org/10.1111/cdev.13575.

Croset, M.-C., and M.-L. Gardes. 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., and M.-L. Gardes. 2020. "Une carte des connaissances pour la construction du nombre en maternelle." *Revue de Mathématiques pour l'École* 233: 117–127.

Darnon, C., and M. Fayol. 2021. "Can an Early Mathematical Intervention Boost the Progress of Children in Kindergarten ? A Field Experiment." *European Journal of Psychology of Education* 3 (1): 1–18. https://doi.org/10.1007/s10212-021-00550-4.

Dong, Y., D. H. Clements, C. A. Day-Hess, J. Sarama, and D. Dumas. 2021. "Measuring Early Childhood Mathematical Cognition : Validating and Equating two Forms of the Research-based Early Mathematics Assessment." *Journal of Psychoeducational Assessment* 39 (8): 983–998. https://doi.org/10.1177/07342829211037195.

Douady, R. 1984. "Jeux de cadre et dialectique outil-objet dans l'enseignement des mathématiques-une réalisation dans tout le cursus primaire." Thèse d'Etat, Université Paris VII.

Douady, R. 1986. "Jeux de cadre et dialectique outil-objet." *Recherches en Didactique des Mathématiques* 7 (2): 2.

Douady, R. 1991. "Tool, Object, Setting, Window : Elements for Analysing and Constructing Didactical Situations in Mathematics." In *Mathematical Knowledge : Its Growth through Teaching*, edited by A. J. Bishop, S. Mellin-Olsen, and J. Van Dormolen, 107–130. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-2195-0_6.

Douady, R. 2003. "Enseignement de la dialectique outil-objet et des jeux de cadres en formation mathématique des professeurs d'école." In *Concertum – Carnets de route de la COPIRELEM*, edited by COPIRELEM, Tome 3, 189–200. Paris: ARPEME.

Douady, R., and M.-J. Perrin-Glorian. 1989. "Un processus d'apprentissage du concept d'aire de surface plane." *Educational Studies in Mathematics* 20 (4): 387–424. https://doi.org/10.1007/BF00315608.

Dowker, A. 2008. "Individual Differences in Numerical Abilities in Preschoolers." *Developmental Science* 11 (5): 650–654. https://doi.org/10.1111/j.1467-7687.2008.00713.x.

Duncan, G. J., C. J. Dowsett, A. Claessens, K. Magnuson, A. C. Huston, P. Klebanov, L. S. Pagani, et al. 2007. "School Readiness and Later Achievement." *Developmental Psychology* 43 (6): 1428. https://doi.org/10.1037/0012-1649.43.6.1428.

Duncan, G. J., and K. Magnuson. 2013. "Investing in Preschool Programs." *Journal of Economic Perspectives* 27 (2): 109–132. https://doi.org/10.1257/jep.27.2.109.

Espy, K. A., M. M. McDiarmid, M. F. Cwik, M. M. Stalets, A. Hamby, and T. E. Senn. 2004. "The Contribution of Executive Functions to Emergent Mathematic Skills in Preschool Children." *Developmental Neuropsychology* 26 (1): 465–486. https://doi.org/10.1207/s15326942dn2601_6.

Fayol, M., and X. Seron. 2005. "About Numerical Representations : Insights from Neuropsychological, Experimental, and Developmental Studies." In *Handbook of Mathematical Cognition*, edited by J. I. D. Campbell, 3–22. New York, NY: Psychology Press. https://doi.org/10.4324/9780203998045.

Floyd, R. G., R. Hojnoski, and J. Key. 2006. "Preliminary Evidence of the Technical Adequacy of the Preschool Numeracy Indicators." *School Psychology Review* 35 (4): 627–644. https://doi.org/10.1080/02796015.2006.12087967.

Fuchs, D., and L. S. Fuchs. 2006. "Introduction to Response to Intervention : What, why, and How Valid Is It?" *Reading Research Quarterly* 41 (1): 93–99. https://doi.org/10.1598/RRQ.41.1.4.

Fuson, K. C. 1988. *Children's Counting and Concepts of Number*. New York: Springer. http://link.springer.com/10.1007978-1-4612-3754-9.

Gardes, M.-L., A. Déglon, S. Javet-Schlegel, C. Turcotte, and M.-C. Croset. 2021. "Analyse des activités proposées dans 'Nombres & opérations' des MER 1-2H." *Revue de Mathématiques pour l'École* 235:39–49.

Gilmore, C., S. Clayton, L. Cragg, C. McKeaveney, V. Simms, and S. Johnson. 2018. "Understanding Arithmetic Concepts : The Role of Domain-Specific and Domain-General Skills." *PLoS ONE* 13 (9): e0201724. https://doi.org/10.1371/journal.pone.0201724.

Ginsburg, H., and A. J. Baroody. 2003. *TEMA-3 : Test of Early Mathematics Ability*. Austin, TX: Pro-ed.

Gormley Jr, W. T., D. Phillips, and T. Gayer. 2008. "Preschool Programs Can Boost School Readiness." *Science* 320 (5884): 1723–1724. https://doi.org/10.1126/science.1156019.

Grégoire, J., M. Noel, and C. Nieuwenhoven. 2003. *Tedi-Math : A Test for Diagnostic Assessment of Mathematical Disabilities*. Belgium: Harcourt Test Publishers.
Grugeon-Allys, B. 2016. "Modéliser le profil diagnostique des élèves dans un domaine mathématique et l'exploiter pour gérer l'hétérogénéité des apprentissages en classe : Une approche didactique multidimensionnelle." *e-JIREF* 2 (2): 2.
Kouki, R., F. Belhaj Amor, and Y. Hachaichi. 2016, mars. Comparaison entre l'évolution historique ayant mené aux développements limités et leur pratique d'enseignement au début de l'université : Entre syntaxe et sémantique. *INDRUM* 2016. https://hal.science/hal-01330197.
Lafay, A., and M. Helloin. 2016. Examath 8-15, batterie informatisée d'examen des habiletés mathématiques. *Grenade: HappyNeuron*.
Li, L., X. Zhou, X. Gao, and D. Tu. 2020. "The Development and Influencing Factors of Kindergarteners' Mathematics Problem Solving Based on Cognitive Diagnosis Assessment." *ZDM Mathematics Education* 52:677–690. https://doi.org/10.1007/s11858-020-01153-x.
Lillard, A. 2012. "Preschool Children's Development in Classic Montessori, Supplemented Montessori, and Conventional Programs." *Journal of School Psychology* 50 (3): 379–401. https://doi.org/10.1016/j.jsp.2012.01.001.
Lillard, A., and N. Else-Quest. 2006. "The Early Years : Evaluating Montessori Education." *Science* 313 (5795): 1893–1894. https://doi.org/10.1126/science.1132362.
Lillard, A., M. Heise, E. Richey, X. Tong, A. Hart, and P. Bray. 2017. "Montessori Preschool Elevates and Equalizes Child Outcomes : A Longitudinal Study." *Frontiers in Psychology* 8:1783. https://doi.org/10.3389/fpsyg.2017.01783.
Lyons, I. M., G. R. Price, A. Vaessen, L. Blomert, and D. Ansari. 2014. "Numerical Predictors of Arithmetic Success in Grades 1–6." *Developmental Science* 17 (5): 714–726. https://doi.org/10.1111/desc.12152.
Margolinas, C., and F. Wozniak. 2012. *Le nombre à l'école maternelle : Une approche didactique*. Bruxelles: De Boeck (Pédagogie et Formation).
Margolinas, C., and F. Wozniak. 2014. "Early Construction of Number as Position with Young Children : A Teaching Experiment." *ZDM* 46 (1): 29–44. https://doi.org/10.1007/s11858-013-0554-y.
Modeste, S. 2020. "Prendre en compte l'épistémologie de l'algorithme. Quels apports d'un modèle de conceptions? Quelle transposition didactique?" *Recherches en didactique des mathématiques* 40 (3): 363–404.
Nguyen, T., T. W. Watts, G. J. Duncan, D. H. Clements, J. S. Sarama, C. Wolfe, and M. E. Spitler. 2016. "Which Preschool Mathematics Competencies Are Most Predictive of Fifth Grade Achievement?" *Early Childhood Research Quarterly* 36:550–560. https://doi.org/10.1016/j.ecresq.2016.02.003.
OECD. 2023. *PISA 2022 Results (Volume I) : The State of Learning and Equity in Education*. OECD. https://doi.org/10.1787/53f23881-en.
Outhwaite, L. A., P. Aunio, J. K. Y. Leung, and J. Van Herwegen. 2024. "Measuring Mathematical Skills in Early Childhood: A Systematic Review of the Psychometric Properties of Early Maths Assessments and Screeners." *Educational Psychology Review* 36 (4): 110. https://doi.org/10.1007/s10648-024-09950-6.
Purpura, D. J., and C. J. Lonigan. 2013. "Informal Numeracy Skills : The Structure and Relations among Numbering, Relations, and Arithmetic Operations in Preschool." *American Educational Research Journal* 50 (1): 178–209. https://doi.org/10.3102/0002831212465332.
Purpura, D. J., and C. J. Lonigan. 2015. "Early Numeracy Assessment : The Development of the Preschool Early Numeracy Scales." *Early Education and Development* 26 (2): 286–313. https://doi.org/10.1080/10409289.2015.991084.
Purpura, D. J., and E. E. Reid. 2016. "Mathematics and Language : Individual and Group Differences in Mathematical Language Skills in Young Children." *Early Childhood Research Quarterly* 36:259–268. https://doi.org/10.1016/j.ecresq.2015.12.020.
Purpura, D. J., E. E. Reid, M. D. Eiland, and A. J. Baroody. 2015. "Using a Brief Preschool Early Numeracy Skills Screener to Identify Young Children with Mathematics Difficulties." *School Psychology Review* 44 (1): 41–59. https://doi.org/10.17105/SPR44-1.41-59.

Riley, M., J. Greeno, and J. Heller. 1983. "Development of Children's Problem-Solving Ability in Arithmetic." In *The Development of Mathematical Thinking*, edited by H. Ginsburg, 153–196. New York, NY: Academic Press.

Sarama, J., and D. H. Clements. 2009. *Early Childhood Mathematics Education Research : Learning Trajectories for Young Children*. New York, NY: Routledge.

Save the Children. 2019. IDELA – The International Development and Early Learning Assessment. Adaptation and Administration Guide. https:// resou rcece ntre. savet hechi ldren. net/ docum ent/idela- the- inter natio nal- devel opment- and- early- learn ing- asses sment/.

Sjoe, N. M., D. Bleses, L. Dybdal, E. Tideman, H. Kirkeby, K. K. Sehested, Hanne Nielsen, Svend Kreiner, and P. Jensen. 2019. "Short Danish Version of the Tools for Early Assessment in Math (TEAM) for 3–6-year-Olds." *Early Education and Development* 30 (2): 238–258. https://doi.org/10.1080/10409289.2018.1544810.

Thomas, A., Y. Tazouti, L. Hoareau, C. Luxembourger, B. Hubert, and A. Jarlégan. 2021. "Early Numeracy Assessment in French Preschool : Structural Analysis and Links with Children's Characteristics." *International Journal of Early Years Education* 31 (4): 1018–1035. https://doi.org/10.1080/09669760.2021.1938518.

Torbeyns, J., C. Gilmore, and L. Verschaffel. 2015. "The Acquisition of Preschool Mathematical Abilities : Theoretical, Methodological and Educational Considerations." *Mathematical Thinking and Learning* 17 (2-3): 99–115. https://doi.org/10.1080/10986065.2015.1016810.

Vergnaud, G. 1982. "A Classification of Cognitive Tasks and Operations of Thought Involved in Addition and Subtraction Problems." In *Addition and Subtraction : A Cognitive Perspective*, edited by T. Carpenter, J. Moser, and T. Romberg, 39–59. Hillsdale, NJ: Routledge.

Watts, T. W., G. J. Duncan, R. S. Siegler, and P. E. Davis-Kean. 2014. "What's Past Is Prologue : Relations between Early Mathematics Knowledge and High School Achievement." *Educational Researcher* 43 (7): 352–360. https://doi.org/10.3102/0013189X14553660.

Weiland, C., C. B. Wolfe, M. D. Hurwitz, D. H. Clements, J. H. Sarama, and H. Yoshikawa. 2012. "Early Mathematics Assessment : Validation of the Short Form of a Prekindergarten and Kindergarten Mathematics Measure." *Educational Psychology* 32 (3): 311–333. https://doi.org/10.1080/01443410.2011.654190.

Woodcock, R. W., K. S. McGrew, and N. Mather. 2001. *Woodcock-Johnson Tests of Achievement*. Itasca, IL: Riverside Publishing.

Wynn, K. 1992. "Addition and Subtraction by Human Infants." *Nature* 358 (6389): 749–750. https://doi.org/10.1038/358749a0.