A Rasch modeling approach for measuring young children's informal mathematics in Peru

Martín Malaspina [1*] (iD), Benito Arias [2] (iD)

[1] Pontificia Universidad Católica del Perú, Lima, PERU
[2] Universidad de Valladolid, Valladolid, SPAIN

Abstract

One of the strongest predictors of future academic achievement is the early and informal math skills children begin their school studies. Because of this, it is essential to have proper tools for measuring the development of mathematical thinking at an early age to be able to intervene in a timelier, more effective way. The purpose of this research is to calibrate the items of informal mathematics from the test of early mathematics ability–third edition (TEMA-3) by applying the Rasch model. A total of 148 Peruvian preschool children (ranging in age from five to six years) participated in the study. The results show good psychometric properties of the informal mathematics dimension of the instrument, which indicates a good fit of the student sample, the items to the proposed model and a tendency toward unbiased items. We further determined that the items analyzed exhibit a consistent internal structure at the theoretical level.

Keywords: early childhood education, informal mathematics, mathematics assessment, psychometrics, Rasch model, a test of early mathematics ability

INTRODUCTION

According to the UNESCO Institute for Statistics (2017), 52% of children and adolescents in Latin America and the Caribbean do not learn the minimum levels of mathematical competence. This percentage is very high, compared to those found in studies of persistent difficulties in learning arithmetic due to neurocognitive factors, such as those by Rapin (2016), where prevalence estimates are between 2% and 7%. One way to explain this significant difference in the percentages of the studies as mentioned above is that most children and adolescents with low mathematical performance have learning difficulties due to environmental or social factors, such as the low quality of education received, few learning opportunities, poor early childhood development, and unfavorable learning environments at home (Fritz et al., 2019).

There is growing evidence and reasons that justify the recent increase in research on early mathematics development (Clements et al., 2017). Some of the main reasons are that several studies (Baroody et al., 2006; Claessens & Engel, 2013; Duncan et al., 2007; Nguyen et al., 2016; Romano et al., 2010; Watts et al., 2014, 2018) have shown evidence that the development of mathematical thinking from an early age is essential to the holistic development of children, forms the basis for various skills and is a significant predictor of future mathematical learning and academic performance.

Given the massive importance of number skill development in early childhood, there has also been a paramount need for adequate measurement tools to help identify children who are having mathematical learning difficulties at this stage of their lives (Clements et al., 2017; Yao et al., 2017). However, there are not many mathematical measurement instruments that are suitable for these early ages (Ryoo et al., 2015). Some examples of instruments that have been used for diagnosis or research with preschool children include the Woodcock-Johnson III tests of achievement (Woodcock et al., 2001), the child math assessment (Klein & Starkey, 2006), the tools for early assessment in mathematics (Clements et al., 2011), the research-based early math assessment (Clements et al., 2008), the number sense core battery (Jordan et al., 2006). Most of these instruments have not been adapted to Spanish and provide a score that attempts to reflect the children's

✉ martin.malaspina@pucp.edu.pe **(*Correspondence)** ✉ benito.arias@uva.es

Contribution to the literature

- This study provides adequate evidence of the validity and reliability for measuring informal mathematics of TEMA-3 (Spanish adaptation) in Peruvian children between the ages of five and six years.
- The methodological contribution of this study is the psychometric analysis of TEMA-3 items using the Rasch model, which overcomes some limitations of the analysis involved using classical test theory.
- The study highlights the importance of the development of informal mathematics and its measurement in early childhood and provides elements for timely diagnoses and appropriate interventions.

overall mathematics ability, which can identify the children's level of mathematical development about peers of the same age or grade (Purpura & Lonigan, 2015).

The test of early mathematics ability–third edition (TEMA-3), developed by Ginsburg and Baroody (2003), is widely used instrument in different countries (e.g., Ryoo et al., 2014; Yao et al., 2017) and has been adapted to Spanish (Ginsburg & Baroody, 2007). It measures children's early mathematics knowledge and helps detect mathematical problems or weaknesses in students. Likewise, due to its design characteristics and its fun, simple application, it is considered one of the preferred options for measuring early arithmetic skills (Nuñez & Pascual, 2011).

TEMA-3 separates the development of children's mathematical abilities into informal and formal. Informal mathematics deals with the notions and processes usually learned in non-school daily dynamics, which arise from interactions with the physical and social environment where scenarios such as games are presented that generate significant learning in a more natural, spontaneous, and enjoyable way (Baroody, 2004; Baroody & Wilkins, 1999; Purpura et al., 2013; Sarama & Clements, 2009). The development of formal mathematics is characterized by the skills and concepts that the child generally learns in school and usually involves more symbolic and written mathematics (Ginsburg & Baroody, 2007). These forms of learning are related to each other to guide the development of mathematical knowledge (Purpura et al., 2013). The National Council of Teachers of Mathematics (NCTM, 2003) complements and highlights the importance of those mentioned above when it states that "the most important connection in early mathematical learning is the one between intuitive, informal mathematics that children have learned through their experiences, and that which they are learning in school." (p. 136).

For early childhood, beginning to understand numbers and operations constitutes informal knowledge (Baroody, 2004), and Sarama and Clements (2009) regard this learning as probably the most important within mathematics. Starkey et al. (2004) complement what was said above by stating that numbering, ordering, comparing numbers, and arithmetic problem solving develop significantly during the preschool years and

form the basis for acquiring formal mathematics in school.

One approach to developing informal mathematics for numbers is that Krajewski and Schneider (2009) proposed, who argue that children go through three overlapping levels according to their evolutionary development. At the first level–basic numerical skills–children learn to distinguish quantities when they compare sets of elements in their environment and at the same time learn the verbal sequence of numbers. Children begin to make meaningful connections at the second level–linking number words with quantity by applying the verbal counting sequence to their respective quantities of set elements, using one-to-one correspondence, and developing the cardinality principle. At the last level–linking quantity relations with number words–children use previously acquired learning to manipulate quantities (basic addition and subtraction) and form new numerical sets.

Children's various ways to count significantly influence the construction of their number concept and their informal learning of mathematics (Baroody, 2004; Baroody & Wilkins, 1999). Gelman and Gallistel (1986) studied counting in depth. They formulated five principles:

1. *One-one principle:* assign a specific number to each element that forms a set,
2. *Stable-order principle:* assign the same numbers to the corresponding elements,
3. *Cardinal principle:* the final number of the series establishes the total number of elements in the set,
4. *Abstraction principle:* the characteristics that the elements possess do not interfere in their counting process, and
5. *Order-irrelevance principle:* the cardinality of the set is the same, regardless of the counting order.

These principles make it clear that children's thinking is not only concrete since as they develop, but they also apply very abstract thinking, such as when applying the principle of one-to-one correspondence or the principle of abstraction, or knowing that adding always implies more, and subtracting implies less (Ginsburg et al., 2008).

Ginsburg and Baroody (2007) highlight two stages of mathematical development related to counting, similar to the levels proposed by Krajewski and Schneider

(2009). The first is pre-counting (nonverbal), which is manifested when young children think about groups of objects and how these vary in terms of quantity. Since they are not yet using words at this age, they likely use mental images. The second stage is counting, which develops as they verbalize numbers using counting words and associate them with quantities.

The National Research Council (NRC, 2009) report on learning mathematics in early childhood identifies the development of numbering, relationships and arithmetic operations as the three main aspects of informal basic mathematics. Likewise, this report states that these three aspects, although related in the broad perspective of the development of early arithmetic, each represents a different domain and is fundamental for the mathematical development of children in early childhood.

In keeping with the above classification, Purpura and Lonigan (2013) note that the development of numbering skills in children implies knowledge about the rules and processes of the counting sequence and the ability to obtain total amounts of elements in a flexible way. Regarding the skills for establishing relationships, they note that these involve how two or more collections or numbers are relevant to each other; in particular, they highlight their association on the mental number line. Furthermore, they give relevance to the verbal and non-verbal aspects of comparing quantities and numbers and the skills necessary to transition between these two aspects. Finally, they specify that arithmetic operations refer to understanding how groups are composed and decomposed by distinguishing sets and subsets.

The NCTM (2000, 2006) proposal differs slightly from this classification, as it integrates the numbering and relationship categories into one and maintains the operations category. Ginsburg and Baroody (2007) agree in various aspects with what were proposed by the NCR (2009) and put forth a similar classification of informal mathematics.

Numbering implies the relationship with only a collection of elements, and, at a basic level, a student should at least know the verbal sequence of numbers. A more advanced concept would imply applying and relating the numerical words with their respective quantities. A higher level would be more advanced and flexible counting techniques, such as correctly saying a sequence of numbers from highest to lowest.

Number comparisons require the presence of two or more collections of elements. This ability is related to number sense and knowledge of how numbers are ordered intuitively. This knowledge helps the child develop the ability to establish "relative distances" between numbers or quantities.

The calculation is introduced when solving simple problems requiring numbers to identify the total number of objects in two sets or the number of objects left when a certain number of its elements are removed. At first, basic counting techniques usually are used, but later these calculations are done mentally. These operations do not require traditional algorithms or writing the calculation procedure.

Concepts are learned knowledge that demonstrates a basic understanding of numerical skills and calculation development in the counting stage. Among these crucial pieces of knowledge are the construction of the cardinal principle, the equivalent distribution, and an understanding of how a whole is related to the parts or elements that comprise it.

In this context, TEMA-3 is valuable because the construction of the items relies on research results of number development and early arithmetic skills. However, despite differentiating between informal mathematics and formal mathematics, no evidence of validity and reliability is shown for the scores of these two dimensions (Ginsburg & Baroody, 2007). Only the analysis of psychometric properties is reported for the total score of early mathematical ability.

Following the above, it is imperative measure the dimension of informal mathematics in children adequately; however, there are no psychometric properties of TEMA-3 with more contemporary measurement models in Latin American contexts. This region is characterized by high levels of poverty and high inequalities in various aspects of people's lives, such as the opportunity gap, gender inequality, and inequalities in access to justice or quality education (Amarante et al., 2022; Balakrishnan et al., 2021; Busso & Messina, 2020). These inequalities have unfortunately increased due to the impact of the COVID-19 pandemic (Acevedo et al., 2022).

Regarding gender inequality in education, existing gaps are known to be a global concern (OECD, 2015). In Latin America, disparities in mathematical learning are known to exist, with boys having an advantage over girls in mathematical performance in most countries (UNESCO, 2018). In relation to these comparisons at preschool age, little research is found, and most of it in contexts other than Latin America. The results of these studies do not show a clear trend, as in some cases there are differences in mathematical skills between girls and boys (e.g., Jordan et al., 2006; Lenes et al., 2022) and in other cases there are not (e.g., Aragón et al., 2013).

In relation to the family context, parents' education is known to be related to their children's learning and development (Lenes et al., 2022). Likewise, Leon and Collahua (2016) show that socioeconomic status is a key variable in explaining academic performance in Peruvian schools. In this context, it is known that maternal education level has been a good predictor of socioeconomic status in studies of early child development and that mothers with better educational backgrounds use richer language with their children and

engage in more complex mathematical interactions at home, which are positive predictors of early numeracy (Lenes et al., 2022; Susperreguy & Davis-Kean, 2016; Susperreguy et al., 2020).

TEMA-3 has been created in the USA and adapted in Spain, countries whose socio-cultural reality is very different from that of Latin America and in particular from that of Peru. Thus, the purpose of this study is to calibrate the informal mathematics items of the Spanish adaptation of TEMA-3 (Ginsburg & Baroody, 2007) by applying the Rasch dichotomous model to the results of young Peruvian children; furthermore, it is analyzed whether the items of the instrument are biased according to the gender of the child and the educational level of the mother. The probabilistic model proposed by Rasch (1960) was chosen because this type of mathematical model has become mainstream as the theoretical basis for measurement. Increasingly, standardized tests are developed from Item Response Theory (IRT) due to having more adequate theoretical measurement principles and its higher potential to solve practical measurement problems (Embretson & Reise, 2000).

METHOD

Participants

Participants in the study were 148 students from the last year of preschool education, 84 girls (56.8%) and 64 boys (43.2%). In order to select these participants, firstly, a convenience sample was chosen, based on the availability of access to public schools in the Metropolitan area of Lima; in total there were six public schools in marginal urban areas. After this, it was decided to include the children who had the informed consent signed by their parents by the agreed date (which was between 30% and 65% of the total number of children per classroom). The children's ages ranged between 62 months and 78 months of age (M=69.2, SD=4.06), and the vast majority of the students spoke Spanish natively (97.2%). The only inclusion criterion was that the children be attending the last year of preschool education in a public school. By contrast, being diagnosed with an intellectual disability was an exclusion criterion for the study.

Regarding the mothers of the children (n=143), they were between 21 and 45 years old (M=31.9, SD=6.0), 33.6% had a paid job, and 32.9% had not graduated from high school (their education level would be from 0 to 3, according to international standard classification of education [ISCED] levels). As for the fathers of the children (n=126), they were between 23 and 56 years old (M=36.0, SD=7.4), 70.9% had a paid job, and 28.4% had not graduated from high school (ISCED levels from zero to three).

Instrument

The TEMA-3 (Ginsburg & Baroody, 2007) is a Spanish adaptation of the 3rd edition of the test published by Ginsburg and Baroody (2003). This standardized test measures the primary mathematics performance of children between the ages of three years and eight years and 11 months of age. The items are constructed based on research results on number development and early arithmetic skills.

The TEMA-3 is administered individually. It is not a timed test, and it has 72 dichotomous (correct or incorrect) items, 41 of which measure four areas of informal mathematics: numbering (e.g., "show me three fingers"), number comparison (e.g., "point to the side with the most points"), the informal calculation (e.g., "Juan has a marble and finds two more, how many does he have in total?") and informal concepts (e.g., "how many stars did you count?"). The other 31 items focus on four areas of formal mathematics: conventions, number facts, formal calculation, and concepts. According to the test manual, the test yields a single value representative of early mathematics ability; however, in one of the few studies that exist on the factor structure of the TEMA-3 (Ginsburg & Baroody, 2003), the authors Ryoo et al. (2015) find that a multidimensional structure fits the data better than a one-dimensional structure.

In this test, the students receive one point for each item answered correctly and zero for wrong answers. The evaluation begins at the point corresponding to the age of the child (for example, those who are five years old start at item 11 and those who are six years old at item 21), and to know the endpoint, it is necessary to take into account each student's ceiling (representing the upper limit of the assessment) and floor (representing the lower limit of the assessment). The ceiling is reached when the child answers five consecutive items incorrectly, and the floor is established using the last five consecutive correct answers. All items below the floor are considered correct.

Psychometric studies of the TEMA-3 (Ginsburg & Baroody, 2007) show adequate reliability results (α>.90, test-retest coefficients>.80). There is also proof of its validity based on the content (basis of the criteria for item construction and selection, analysis of the items from the classical test theory) and construct (differences by age and differentiation of low-performing groups). In addition, the actual test (Ginsburg & Baroody, 2003) exhibits evidence of concurrent validity, correlation coefficients between .54 and .91 with the following tests: Keymath-Revised: A diagnostic inventory of essential mathematics–normative update (Keymath–R/UN; Connolly, 1998), Woodcock-Johnson III tests of achievement (WJ-III ACH; Woodcock et al., 2001), diagnostic achievement battery–third edition (DAB-3; Newcomer, 2001) and young children´s achievement test (YCAT; Hresko et al., 2000). Since most of the

Table 1. Informal mathematics items selected from TEMA-3

Categories of informal mathematics	Items
Numbering	M2, M3, M4, M5, M6, M9, M10, M12, M13, M20, M21, M22, M25, M27, M29, M32, M33, M37
Number comparisons	M1, M16, M17, M26, M35
Calculation	M8, M19, M23, M24, M34
Concepts	M7, M11

psychometric analyses of TEMA-3 have been conducted on samples from the United States, Yao et al. (2017) argue that, given the widespread use of the test, more studies are needed in multiple countries to examine the validity and reliability of the measurement instrument.

Procedure

Pontificia Universidad Católica del Perú's Ethical Research Committee followed and approved all applicable ethical considerations. Meetings were held with principals and teachers of preschool education in their respective schools, where they were given an explanation of the study and asked for their collaboration. Then, the parents of the students were contacted using a written document to obtain permission and informed consent for their children to participate. This document also contained a sociodemographic survey to collect more information on the children and their parents.

Once the signed consent was received, the dates for administering the test of early mathematical ability to the students were arranged with the teachers. At the specified date and time, in a room set up beforehand inside the school, the students were asked to provide informed consent orally and, if affirmative, the TEMA-3 was administered, a process that lasted approximately 20 minutes.

Data Analysis

For our research, we employed the Rasch dichotomous model (1960) implemented in the WINSTEPS program, v. 3.81 (Linacre, 2014, 2019). This model assumes that the trait to be measured can be represented uni-dimensionally, and it considers the items and the persons evaluated together. It is also important to consider the difficulty of each test item and the general skill level of the test-takers to determine the probability that the answer is correct. The author of this model used the following log function to posit a relationship:

$$ln\left(\frac{P_{ni}}{1-P_{ni}}\right) = (\theta_n - \beta_i). \quad (1)$$

In Eq. (1), P_{ni} is the probability that person n answers item i correctly, θ_n is the ability of person n with the latent variable, and β_i is the difficulty of item i.

RESULTS

Since the TEMA-3 has a format in which not all children answer the same number of items, those items that were not answered or were only answered by very few children–fewer than 10–were not selected (Yao et al., 2017). The total number of items considered for the analysis was the first 30 informal mathematics items in the test (**Table 1**).

Before analyzing the model's results in greater detail, it is essential to verify that the data collected satisfied the assumptions specified in the Rasch model. Notable among the results is that the variance explained by the measures is greater than 63%, and the difference between the variance observed and that expected by the model is minimal (0.4%). In addition, since the first contrast has an eigenvalue below three, the test data have a one-dimensional trend.

The items were found to have average reliability of .99, which for the persons it was .92. These are very suitable values because they are close to one (Wright & Stone, 1979). Similarly, the separation index of 9.1 is quite appropriate for the items (it differentiates between 9 different levels of informal mathematics). In the case of the children evaluated, the separation index is 3.5, which indicates that the test differentiates the sample into at least three levels of informal mathematics.

Table 2 shows the results of the Rasch analysis, ordered according to the magnitude of the estimated parameters. The polarity of the items indicates that the point-biserial (PB) correlations range from 0.29 to 0.72, thus confirming that all the items are aligned in the same direction as the latent variable. Moreover, the mean fit and the standard deviations of the items are adequate (infit=0.97; SD=0.26; outfit=0.80; SD=0.62). Similarly, the mean fit, and standard deviations of the individuals are acceptable (infit=0.97; SD=0.45; outfit=0.74; SD=1.02). These results suggest that this set of items fulfils, in principle, the requirements to estimate the children's levels of informal mathematics.

A graphical representation of the MNSQ Infit index about the item difficulty measures is shown in **Figure 1**. There is variety in the distribution of the item measurement, with item 34 and item 1 being the most difficult and the easiest, respectively. In addition, the infit MNSQ values are close to 1, almost all the items are located in the 0.5 to 1.5 zone, or zone of acceptable and productive fit for the measure (Linacre, 2019), with the only exception being item 9, whose Infit MNSQ value showed a minimal deviation from the ideal range (1.62). As for the students, only 11.49% exhibited Infit MNSQ values higher than 1.5. Consequently, the model adjusts well to 88.51% of the children.

Table 2. Estimates of the item parameters

Item	Description	M	SE	Infit		Outfit		PTBISERL-EX		Exact match	
				MNSQ	ZEMP	MNSQ	ZEMP	Corr	Exp	Obs%	Exp%
M34	Count from the highest addend	8.33	1.08	1.16	0.50	0.78	0.30	0.50	0.54	93.80	93.70
M35	Mental number line (2 digits)	6.85	0.73	0.87	-0.30	0.72	0.10	0.60	0.55	84.60	78.90
M37	Count backwards (from 20)	6.70	0.76	1.19	0.70	0.95	0.20	0.49	0.54	60.00	73.30
M33	Count by 10s (up to 90)	6.67	0.62	0.69	-0.80	0.24	-0.50	0.69	0.58	91.90	90.10
M32	Next number (transition by tens to 50)	5.38	0.43	0.68	-1.40	0.64	-0.40	0.68	0.58	91.80	84.90
M29	Count out loud (up to 42)	5.07	0.38	0.89	-0.40	1.63	1.10	0.59	0.58	91.70	86.90
M26	Mental number line (1 digit)	3.50	0.28	1.23	1.50	1.22	0.60	0.52	0.59	75.80	81.80
M21	Next number (two digits, up to 40)	3.29	0.27	0.70	-2.30	0.41	-1.50	0.70	0.60	88.30	82.90
M27	Produce sets (19 elements)	3.27	0.27	1.01	0.10	1.20	0.60	0.58	0.59	82.80	81.00
M24	Mental addition (sum from 5 to 9)	2.31	0.25	0.99	0.00	0.73	-0.60	0.62	0.60	74.00	78.40
M25	Count backwards (from 10)	2.30	0.26	0.97	-0.20	0.77	-0.50	0.62	0.60	79.60	78.20
M17	Number comparison (from 5 to 10)	2.07	0.24	0.91	-0.70	1.74	1.50	0.64	0.64	84.00	81.30
M19	Addition word problems (concrete objects)	2.05	0.25	1.27	2.10	1.41	1.00	0.55	0.63	74.20	79.90
M23	Addition word problems (modeling)	0.88	0.26	1.28	2.10	2.80	2.90	0.45	0.57	74.00	79.40
M20	Count out loud (up to 21)	0.79	0.25	1.00	0.00	0.75	-0.50	0.62	0.60	80.50	79.70
M13	Next number (from 1 to 9)	0.34	0.24	0.76	-2.00	1.38	0.90	0.72	0.69	89.80	83.80
M16	Number comparison (from 1 to 5)	-0.23	0.26	0.87	-0.90	0.59	-0.80	0.68	0.65	89.00	84.30
M22	Counting (from 6 to 10)	-1.80	0.42	1.42	1.50	1.77	1.10	0.32	0.45	91.40	92.70
M11	Number constancy	-2.39	0.31	1.04	0.30	0.77	-0.20	0.64	0.65	89.80	90.30
M12	Create sets (up to 5 elements)	-2.39	0.31	1.13	0.70	0.84	0.00	0.63	0.65	91.20	90.30
M10	Show fingers (up to 5)	-3.13	0.34	1.21	1.00	0.61	-0.40	0.59	0.61	89.10	92.00
M8	Non-verbal addition and subtraction with objects	-3.25	0.35	0.94	-0.20	0.38	-0.90	0.62	0.61	93.20	92.30
M7	Cardinality rule	-3.93	0.39	0.54	-2.30	0.13	-1.80	0.63	0.57	96.60	93.90
M6	Counting (from 1 to 5)	-4.25	0.41	0.50	-2.40	0.11	-1.90	0.62	0.55	96.60	94.50
M9	Count by 1 (up to 10)	-4.80	0.45	1.62	1.90	0.80	0.00	0.44	0.51	93.20	95.50
M5	Non-verbal production (from 1 to 4 elements)	-6.04	0.55	1.11	0.40	0.24	-1.50	0.41	0.41	96.60	97.00
M3	Intuitive number sense	-6.36	0.58	0.52	-1.60	0.05	-2.80	0.44	0.38	98.60	97.20
M4	Count by 1 (up to 5)	-6.36	0.58	1.04	0.20	0.13	-2.10	0.39	0.38	97.30	97.20
M2	Show fingers (1, 2, many)	-7.16	0.69	0.78	-0.50	0.06	-2.90	0.34	0.31	98.60	98.00
M1	Perception of more (up to 10 elements)	-7.71	0.80	0.72	-0.40	0.04	-3.30	0.29	0.25	98.60	98.60
Items	Mean	0.00	0.43	0.97	-0.10	0.80	-0.40			87.90	87.60
	SD	4.60	0.21	0.26	1.20	0.62	1.40			9.10	7.30
Persons	Mean	0.63	0.77	0.97	-0.10	0.74	0.00			89.30	88.60
	SD	3.11	0.12	0.45	1.00	1.02	0.60			7.10	3.00

Note. M: Estimate (calibration) of the parameter; SE: Standard error of the estimate; Infit & Outfit MNSQ: Weighted mean-square standardized information statistics; PTBISERL-EX (Corr & Exp): Point-biserial correlation (observed and expected by the model); Exact match (Obs% & Exp%): Percentage of points that fit the prediction and percentage expected by the model
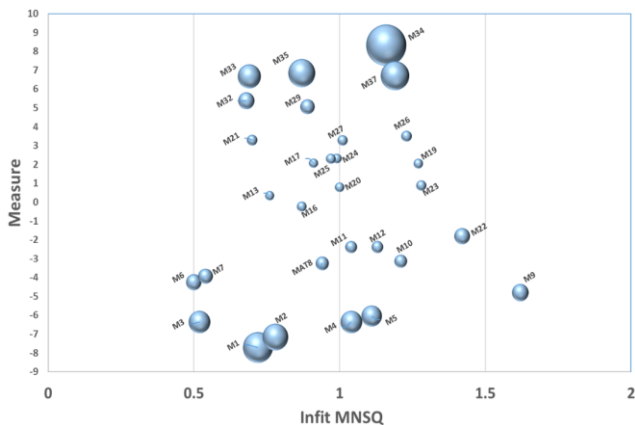

Figure 1. Distribution of items by difficulty & infit MNSQ (Diameter of bubbles corresponds to standard error)

To graph how the difficulty of the items is related to the levels of the children in the latent trait evaluated (informal mathematics), maps of persons and items, called "Wright maps" (**Figure 2**), are used.

Since the Rasch model uses the logit measure, which is the same unit of measure for the difficulty of the items and the children's ability, both metrics can be compared on the same graph to analyze, thus if the difficulty of the items is or is not suitable for the children evaluated. **Figure 2** shows a considerable overlap or alignment between the two measures (*targeting*), and their distributions fluctuate between approximately -8 and 9 logit. Moreover, the difficulty measures of the items exhibit a uniform trend, and the children's ability a usual trend. All these characteristics indicate that the difficulty of the items is adequate for the sample selected and that
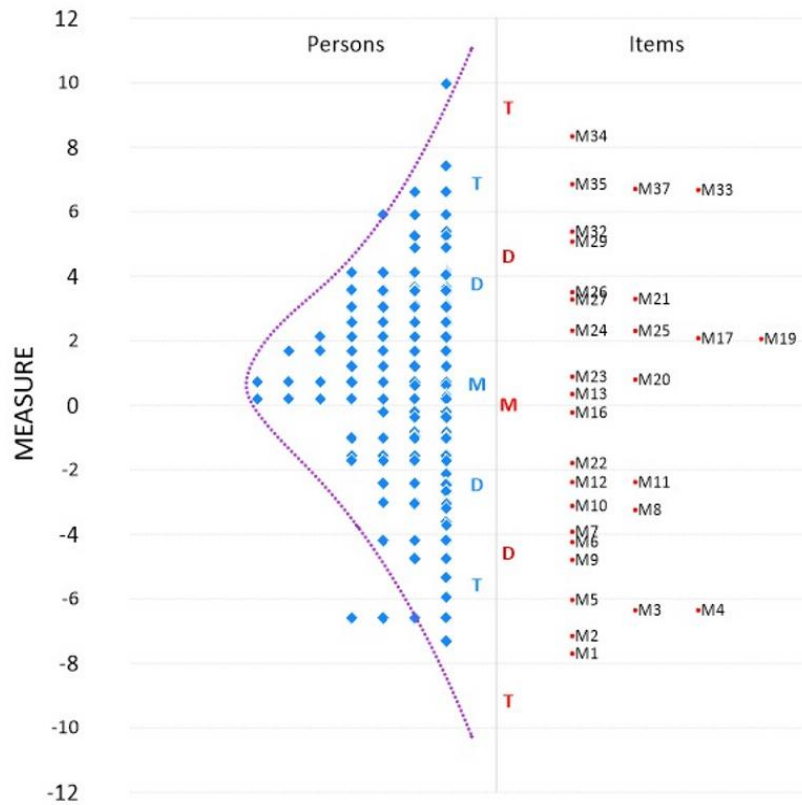
Figure 2. Wright map

the 30 items can evaluate and differentiate between the different levels of informal mathematics of the children.

The TEMA-3 items were organized from easy to difficult using data from the adapted test with a Spanish sample (Ginsburg & Baroody, 2007). To determine whether the order found based on the difficulty of the items in the Peruvian sample was consistent with the order proposed in the Spanish adaptation of the test, the Kendall tau-b correlation coefficient was calculated (Kendall & Gibbons, 1990). We found a positive, high, and statistically significant correlation between the order of the items in the samples from Spain and Peru (τ_b=0.89, p<.001).

As for the accuracy of the measurement, we estimated the test information function. This revealed a distribution with a bimodal trend, which means that, in two intervals, the informal mathematical test measures with greater accuracy the first interval for values of theta between approximately θ=-4 and θ=-2, and the second interval for values of theta between approximately θ=1 and θ=3. The highest standard measurement errors are found at the ends of the continuum.

The differential item functioning (DIF) analysis seeks to identify if the items of an instrument are biased in some way. This can be done by comparing whether the items work similarly for two or more different groups. The uniform DIF as a function of the participants' gender (**Figure 3**) shows that for most items no statistically significant differences are found, which is a good indicator of test invariance. However, item 9 (count by

ones to 10) is 2.82 logits more difficult for the boys (t(134)=3.05; p=.003) than for the girls, while item 12 (create sets, up to five elements) is 1.51 logits more difficult for the girls (t(125)=-2.05; p=.042) than for the boys.

We also conducted a uniform DIF analysis based on the educational level of the participants' mothers (**Figure 4**). Most of the items do not show DIF according to whether or not the mothers graduated. We only found that item 4 (count by ones to five) has a particular risk of exhibiting DIF since it is 3.56 logits more difficult for participants whose mothers did not finish school (t(136)=1.76; p=.080) than for children whose mothers did finish school.

DISCUSSION

There is a growing interest in studying the early number skills and informal learning associated with children's mathematical development. From this perspective, our research contributes to this field of study by reviewing the theoretical foundations and analyzing the psychometric properties and calibration of TEMA-3 items involving informal mathematics for a sample of Peruvian preschool education students between five and six years of age. To the best of our knowledge, this study is the first to calibrate TEMA-3 items using the Rasch dichotomous model in a Peruvian sample; furthermore, the specific psychometric analysis of the informal mathematics aspect of the test has received little attention worldwide.
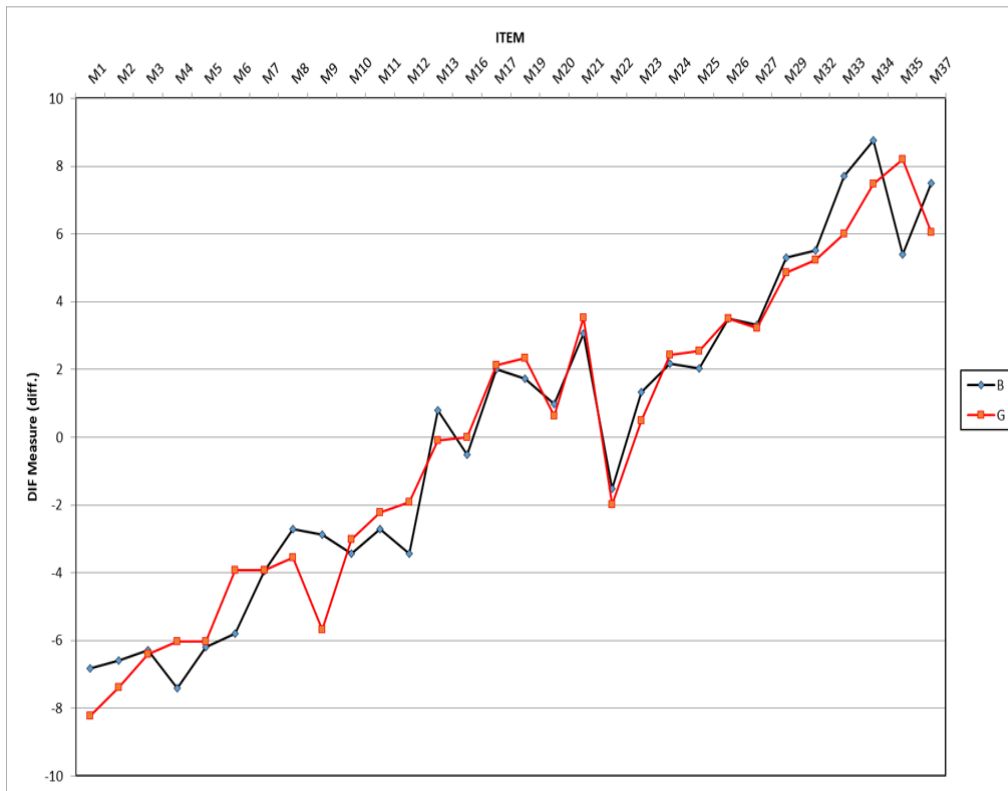
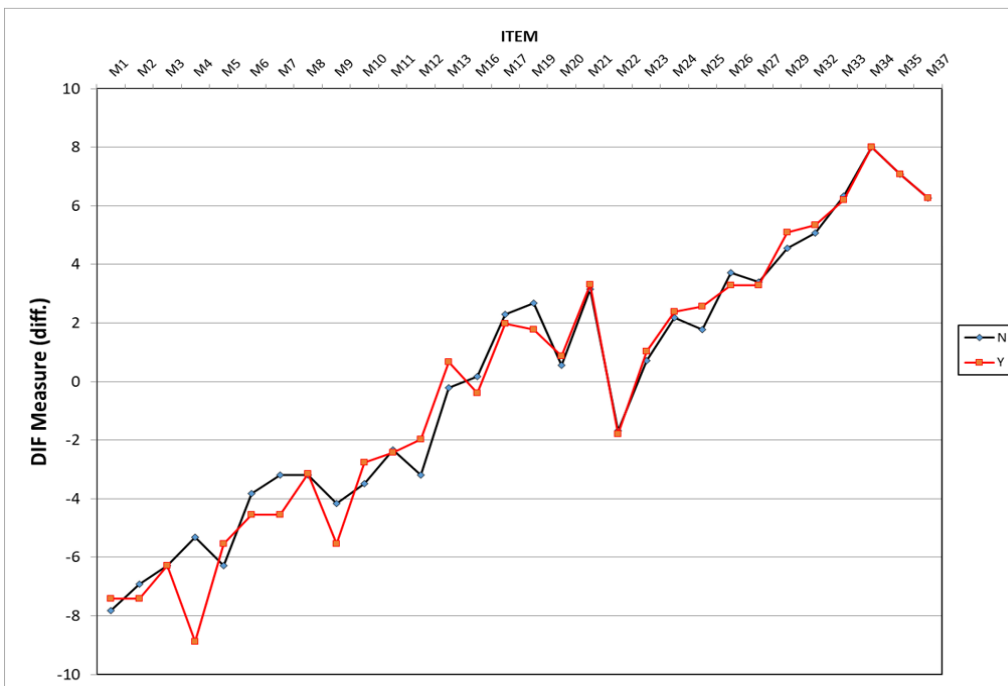**Figure 3.** DIF values by gender (B: Boys & G: Girls)

Figure 4. DIF values by the educational level of the mother (N: Group of children with mothers who did not finish school & Y: Group of children with mothers who did finish school)

Our Rasch model analysis results indicate that the informal mathematics items of TEMA-3 generally exhibit good technical qualities (overall fit and reliability) and an adequate internal structure (consistency between the results and the theoretical design of the test items). To analyze these general results in greater detail, it is pertinent to ask whether the results obtained from the informal mathematics of TEMA-3 comply with the theoretical aspects of informal number skills. To answer this question, we have to consider how well the data fit the Rasch model, which involves the overall fit, the fit of the items, and the individuals' fit. It is also essential to examine the consistency at the theoretical level of the estimated order of the difficulties

of the items and compare it with the proposed order of the items in the design of the test used.

After verifying the assumptions, we analyzed the overall fit of the data to the proposed model. In general, we found that the items selected from the TEMA-3 encompass a wide range of informal mathematical skills with a high overall reliability index. This shows a good distribution of items along the continuum of the latent variable. We also found that the average reliability index of the individuals is high, which suggests that the items differentiate the children in the sample well in terms of their level of informal mathematics.

A more detailed analysis of the fit of the items reveals that item 9 (recite the number sequence from 1 to 10 while the evaluator points to the objects) is the only one that is slightly out of fit based on the Infit MNSQ statistic. This could be due to problems with the item's content, or that it is not measuring a single latent variable, or that its discrimination is dissimilar from that of the other items. Since this item only measures the memorization of the first 10 numbers, there may be a high load from another latent trait of the individual. However, since it is a minimal outfit that does not distort or degrade the measure or the construct (Linacre, 2002), it would not be necessary to eliminate this item from the test.

An item that is out of fit using the Outfit MNSQ statistic and that could indicate a lack of homogeneity with other test items (Linacre, 2002) is item 23 (oral addition problems with modelling, e.g., "Juan has six marbles and gets two more, how many does he have in total?"). The mismatch could be caused, in part, by the relationship between the item with the various levels of language development and oral comprehension that children have at this age and in more disadvantaged social contexts, such as those of the participants whose parents have low level education, which could distort the measurement of a single latent variable using this item. However, since the Infit MNSQ has a good fit, which is less sensitive to outliers than the outfit (Linacre, 2002), it might be acceptable to include the item in the test, but with caution. As a result, the great majority of the data found using the 30 TEMA-3 items can be adequately explained by the Rasch dichotomous model.

Regarding the detail of the fit of the persons, we found that the model adequately explains the answers given by 88.51% of the participants evaluated, which indicates that the Rasch model cannot be used to adequately explain the answer patterns of only 11.49% of the students evaluated. It is thus possible to state that the test helps measure informal mathematics in a sample or population with characteristics similar to that selected for this study, which is very valuable from applying the instrument in the context of Peruvian education.

To finish answering the initial question regarding the internal structure of the test, we proceeded to analyze other properties of the items that can be discerned by applying the Rasch model. If we observe the position of the items along the continuum of the latent variable measured (for example, with the Wright map), we can analyze two main aspects:

1. if the items are ordered homogeneously and
2. if the hierarchical position coincides with the theoretical order proposed by the test.

The items were distributed over a wide range and do not contain any noticeable gaps regarding the first point. One range where there could be a significant gap is between item 22 (-1.80 logits) and 16 (-0.23 logits). However, the staggering along the general measurement interval tends to be adequate, so it seems unnecessary to add items to fill in some small information gaps.

Regarding the second point, it should be noted that, in general, the observed structure of the difficulty estimates of the test items is consistent with the developmental progression of informal mathematics in children. Precisely, the items located in the lower part of the Wright map mainly assess the primary number and pre-counting skills (e.g., item 2: show specific numbers of fingers on one hand, item 9: recite the number sequence from 1 to 10); items located in the center assess counting skills in general and link number words to quantities (e.g., item 16: identify the larger of two digits, from 1 to 5, only numbers are mentioned and no reference is made to objects, item 22: count sets of up to 10 randomly arranged dots); and the items located at the top assess more advanced counting skills, which link quantitative relationships with number words (e.g., item 27: separate 19 objects from a larger set, item 34: addition problems that are solved by counting from the larger addend) (Ginsburg & Baroody, 2007; Krajewski & Schneider, 2009).

Continuing with the analysis of the relationship between the empirical order of the difficulties of the items and the theoretical aspects of informal mathematics, we see that the items with the lowest level of difficulty are usually those that are accompanied by external representations, such as specific images or objects (e.g., item 4: counting fingers); that medium-level item typically use a combination of external representations and verbal instruction (e.g., item 23: solving oral addition problems with tokens); and that more complex items are mainly based on verbal instruction (e.g., item 32: the student is asked what number comes after 29, then the question is repeated with the number 49). These findings are consistent with the evolutionary development of children since they initially tend to think and carry out tasks with more concrete objects, and then they incorporate counting principles that allow them to carry out tasks in a more abstract way (Gelman & Gallistel, 1986; Ginsburg et al., 2008).

To finish the analysis of the second point, it is essential to compare whether the empirical order found

with the difficulty of the items in the Peruvian sample is consistent with the order proposed in the design of the test used, which is a Spanish adaptation. This is highly relevant since the test uses the floor and ceiling rule for the application, which implies that the order of the items must be meticulously planned since if it were not, it would affect the measurement of the student's mathematical abilities. Graphically, the Wright Map shows a general match between the patterns of the order followed in both samples (Spanish and Peruvian); however, to define this observation, the correlation between these rankings was calculated and found to be positive, high, and statistically significant, which indicates that the ranking is very similar and that it is consistent with the theory. The minimal differences found could be attributed to the fact that the order of the items in the Spanish adaptation was estimated using a two-parameter IRT model, and also that the sample was made up of children between three and eight years of age (Ginsburg & Baroody, 2007).

To have information on the bias of each item, we resorted to a DIF analysis involving the child's gender and the mother's educational level to segment groups and look for significant differences between them. Regarding gender, a potential DIF was found in item 9 (reciting the number sequence from 1 to 10), which was easier for girls than for boys; however, it would be worthwhile to check if the DIF is also present in other samples. We then conducted a DIF analysis involving the educational level of the participants' mothers. We found no DIF with any item, which would indicate that the test items, in general, tend to be unbiased. These are some initial assessments using DIF, but it would be beneficial to do it with other significant variables in the construct. For example, as Yao et al. (2017) mentioned, more studies are needed to analyze whether the TEMA-3 items work invariably in different countries. They could be used to provide evidence of validity through cross-cultural comparisons.

All the results analyzed in the above paragraphs show that this psychometric and methodological approach offers several benefits and advantages over the traditional approach of classical test theory. The successful application of the Rasch model to the informal mathematics items of TEMA-3 empirically demonstrates certain advantages; for example, the joint measurement of items and persons, the direct comparison between the skills of the persons and the difficulties of the items, the confirmation that the difficulty of the items is not the same in a test, the application of contrasts to corroborate the invariance of the item parameters, the specificity of the standard measurement error and the amount of information used to measure each point of informal mathematics.

The limitations of this research include the following. First, due to the sample size and the intentional way in which participants were selected, the results cannot be generalized. However, the methodology presented in detail provides sufficient elements for the study to be replicated with larger samples and in different social realities. Another aspect to consider is that the informal mathematics items of TEMA-3 were calibrated only with data from children ages five to six, which caused the loss of information on the items contained at the beginning of the test due to the rules on the floor of the test, items were assumed as correct in many cases. As a third limitation, we might mention that the size and diversity of the sample were not very large, which may be a weakness when analyzing the DIF. In future research, samples from private educational institutions and rural areas could be included to have a more representative sample of Peru and more subgroups for comparisons. Finally, this research focused on aspects involving numbers and operations in informal mathematics; however, it would also be appropriate to include other dimensions of young children's mathematical skills, such as algebra, patterns, geometry, statistics, and probability.

This study provides a valid way to measure informal mathematics in young children, which is often invisible in teaching practice. The information obtained through this assessment could facilitate appropriate interventions at an early age or further research to shed more light on the relationship between informal and formal mathematics. Moreover, by virtue of the application of the test and the interpretation of the results, principals and teachers of early childhood education can become more aware of the importance of informal mathematics as a generator of children's learning potential, which is often neglected.

We hope that our results can contribute to new research on informal mathematics and TEMA-3 and the development of new measurement instruments. For example, more items could be proposed that more adequately address more current and complementary theoretical proposals, such as those mentioned by the NRC (2009) and those of Purpura and Lonigan (2013). Reflecting on these results and motivating the generation of innovative proposals for intervention strategies in early childhood mathematics education would provide more and better elements for a forward-looking education, especially in the Peruvian context.

REFERENCES

Acevedo, I., Castellani, F., Cota, M. J., Lotti, G., & Székely, M. (2022). *El aumento de la desigualdad en*

América Latina: Un efecto colateral de la pandemia [*The increase in inequality in Latin America: A collateral effect of pandemic*]. Banco Interamericano de Desarrollo. https://doi.org/10.18235/0003967

Amarante, V., Colacce, M., & Scalese, F. (2022). Poverty and gender in Latin America: How far can income-based measures go? *Journal of International Development, 34*, 109-129. https://doi.org/10.1002/jid.3584

Aragón, E. L., Delgado, C. I., Aguilar, M., Araújo, A., & Navarro, J. I. (2013). Estudio de la influencia de la inteligencia y el género en la evaluación matemática temprana [Study of the influence of intelligence and gender in the assessment of early math]. *European Journal of Education and Psychology, 6*(1), 5-18. https://doi.org/10.30552/ejep.v6i1.90

Balakrishnan, R., Lizarazo, S., Santoro, M., Toscani, F. G., & Vargas, M. (2021). *Commodity cycles, inequality, and poverty in Latin America.* International Monetary Fund.

Baroody, A. J. (2004). The developmental bases for early childhood number and operations standards. In D. H. Clements, J. Sarama, A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 173-219). Lawrence Erlbaum Associates Publishers.

Baroody, A. J., & Wilkins, J. L. (1999). The development of informal counting, number, and arithmetic skills and concepts. In J. V. Copley (Ed.), *Mathematics in the early years* (pp. 48-65). National Council of Teachers of Mathematics.

Baroody, A. J., Lai, M.-l., & Mix, K. S. (2006). The development of young children's early number and operation sense and its implications for early childhood education. In B. Spodek, & O. N. Saracho (Eds.), *Handbook of research on the education of young children* (pp. 187-221). Lawrence Erlbaum Associates Publishers.

Busso, M., & Messina, J. (2020). *La crisis de la desigualdad: América Latina y el Caribe en la encrucijada* [*The crisis of inequality: Latin America and the Caribbean at the crossroads*]. Banco Interamericano de Desarrollo. https://doi.org/10.18235/0002629

Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record, 115*(6), 1-29. https://doi.org/10.1177/016146811311500603

Clements, D. H., Sarama, J., & Liu, X. (2008). Development of a measure of early mathematics achievement using the Rasch model: The research-based early math assessment. *Educational Psychology, 28*(4), 457-482. https://doi.org/10.1080/01443410701777272

Clements, D. H., Sarama, J., & Wolfe, C. B. (2011). *TEAM–Tools for early assessment in mathematics.* McGraw-Hill Education.

Clements, D. H., Sarama, J., Day-Hess, C., Germeroth, C., Ganzar, J., Pugia, A., & Barker, J. (2017). *Comprehensive review of assessments of early childhood mathematics competencies*[Unpublished manuscript].

Connolly, A. J. (1998). *Manual for the KeyMath revised/NU: A diagnostic inventory of essential mathematics.* American Guidance Service

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428-1446. https://doi.org/10.1037/0012-1649.43.6.1428

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates Publishers.

Fritz, A., Geraldi, V., & Räsänen, P. (2019). Introduction. In A. Fritz, V. Geraldi, & P. Räsänen (Eds.), *International handbook of mathematical learning difficulties* (pp. 1-8). Springer. https://doi.org/10.1007/978-3-319-97148-3_1

Gelman, R., & Gallistel, C. R. (1986). *The child's understanding of number*. Harvard University Press. https://doi.org/10.4159/9780674037533

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early mathematics ability*. PRO-ED.

Ginsburg, H. P., & Baroody, A. J. (2007). *TEMA-3: Test de competencia matemática básica* [*TEMA-3: Test of basic mathematics competence*]. TEA Ediciones.

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: What it is and how to promote it. *Social Policy Report, 22*(1), 3-22. https://doi.org/10.1002/j.2379-3988.2008.tb00054.x

Hresko, W. P., Peak, P. K., Herron, S. R., & Bridges, D. L. (2000). *Young children's achievement test*. PRO-ED.

Jordan, N. C., Kaplan, D., Olah, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*(1), 153-175. https://doi.org/10.1111/j.1467-8624.2006.00862.x

Kendall, M. G., & Gibbons, J. D. (1990). *Rank correlation methods*. Griffin.

Klein, A., & Starkey, P. (2006). *Child math assessment*. University of California, Berkeley.

Krajewski, K., & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a

four-year longitudinal study. *Learning and Instruction, 19*(6), 513-526. https://doi.org/10.1016/j.learninstruc.2008.10.002

Lenes, R., Størksen, I., McClelland, M. M., & Idsøe, T. (2022). The role of mother's education and child gender for children's vocabulary and math skills in the transition from early childhood education and care to first grade in Norway. *European Early Childhood Education Research Journal, 30*(3), 403-422. https://doi.org/10.1080/1350293X.2022.2055101

León, J., & Collahua, Y. (2016). El efecto del nivel socioeconómico en el rendimiento de los estudiantes peruanos: Un balance de los últimos 15 años [The effect of the socioeconomic level on the performance of Peruvian students: A balance of the last 15 years]. In GRADE. *Investigación para el desarrollo en el Perú: Once balances* [*Research for development in Peru: Eleven balances*] (pp. 109-162). GRADE.

Linacre J. M. (2014). *Winsteps Rasch measurement* [Computer software]. Winsteps.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*, 878.

Linacre, J. M. (2019). *A user's guide to WINSTEPS* [Computer manual]. Winsteps. https://www.winsteps.com/manuals.htm

NCTM. (2000). *Principles and standards for school mathematics*. National Council of Teachers of Mathematics.

NCTM. (2003). *Principios y estándares para la educación matemática* [*Principles and standards for mathematical education*]. Sociedad Andaluza de Educación Matemática Thales [Andalusian Mathematical Education Society Thales].

NCTM. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics*. National Council of Teachers of Mathematics.

Newcomer, P. L. (2001). *Diagnostic achievement battery*. PRO-ED.

Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (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

NRC. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. The National Academies Press. https://doi.org/10.17226/12519

Nuñez, M. C. & Pascual, M. I. (2011). Habilidades matemáticas básicas en alumnos de 3º de Infantil: Detección temprana de dificultades de aprendizaje y orientaciones para la intervención [Basic mathematical skills in students of 3rd Infant: early detection of learning difficulties and orientations for intervention]. *Revista Diálogo Educacional* [*Educational Dialogue Magazine*], *11*(32), 83-105. https://doi.org/10.7213/rde.v11i32.4087

OECD. (2015). *The ABC of gender equality in education: Aptitude, behaviour, confidence*. PISA, OECD. https://doi.org/10.1787/9789264229945-en

Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Educational Psychology, 105*(2), 453-464. https://doi.org/10.1037/a0031753

Purpura, D. J., & Lonigan, C. J. (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., & Lonigan, C. J. (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

Rapin, I. (2016). Dyscalculia and the calculating brain. *Pediatric Neurology, 61*, 11-20. https://doi.org/10.1016/j.pediatrneurol.2016.02.007

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research.

Romano, E., Babchishin, L., Pagani, L. S., & Kohen, D. (2010). School readiness and later achievement: Replication and extension using a nationwide Canadian survey. *Developmental Psychology*, *46*(5), 995-1007. https://doi.org/10.1037/a0018880

Ryoo, J. H., Molfese, V. J., Brown, E. T., Karp, K. S., Welch, G. W., & Bovaird, J. A. (2015). Examining factor structures on the test of early mathematics ability–3: A longitudinal approach. *Learning and Individual Differences*, *41*, 21-29. https://doi.org/10.1016/j.lindif.2015.06.003

Ryoo, J. H., Molfese, V. J., Heaton, R., Zhou, X., Brown, E. T., Prokasky, A., & Davis, E. (2014). Early mathematics skills from prekindergarten to first grade: Score changes and ability group differences in Kentucky, Nebraska, and Shanghai samples. *Journal of Advanced Academics, 25*(3), 162-188. https://doi.org/10.1177/1932202X14538975

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. Routledge. https://doi.org/10.4324/9780203883785

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early*

Childhood Research Quarterly, 19(1), 99-120. https://doi.org/10.1016/j.ecresq.2004.01.002

Susperreguy, M. I., & Davis-Kean, P. E. (2016). Maternal math talk in the home and math skills in preschool children. *Early Education and Development, 27*(6), 841-857. https://doi.org/10.1080/10409289.2016.1148480

Susperreguy, M. I., Douglas, H., Xu, C., Molina-Rojas, N., & LeFevre, J. A. (2020). Expanding the home numeracy model to Chilean children: Relations among parental expectations, attitudes, activities, and children's mathematical outcomes. *Early Childhood Research Quarterly, 50*, 16-28. https://doi.org/10.1016/j.ecresq.2018.06.010

UNESCO Institute for Statistics. (2017). *More than one-half of children and adolescents are not learning worldwide.* UNESCO Institute for Statistics.

UNESCO. (2018). *Achieving gender equality in education: Don't forget the boys* [Policy paper]. UNESCO. https://en.unesco.org/gem-report/node/2426

Watts, T. W., Duncan, G. J., Clements, D. H., & Sarama, J. (2018). What is the long-run impact of learning mathematics during preschool? *Child Development, 89*(2), 539-555. https://doi.org/10.1111/cdev.12713

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*, 352-360. https://doi.org/10.3102/0013189X14553660

Woodcock, R., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson tests of achievement.* Riverside Publishing.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement.* MESA Press.

Yao, S. Y., Muñez, D., Bull, R., Lee, K., Khng, K. H., & Poon, K. (2017). Rasch modeling of the test of early mathematics ability–Third edition with a sample of K1 children in Singapore. *Journal of Psychoeducational Assessment, 35*(6), 615-627. https://doi.org/10.1177/0734282916651021

https://www.ejmste.com