

# Validity and Reliability of Mathematics Research Instruments: Systematic Review

Fani Esperanza Mejía-Clavo<sup>1</sup>, Oscar López- Regalado<sup>1</sup>,  
Jessica del Rocío Fernández Velásquez<sup>1</sup>,  
Antony Esmít Franco Fernández Altamirano<sup>1</sup>, Hilda Aurora Torres Cueva<sup>1</sup>

<sup>1</sup> Graduate School, Cesar Vallejo University, Chiclayo, Peru

**Abstract** - The difficulties students face in mathematics are due to various factors that hinder their logical understanding. It is crucial that teachers abandon traditional teaching models and adopt new methodologies to overcome these challenges. However, effective implementation of these methodologies requires valid and reliable assessment tools, which has been little explored in previous studies. The objective of this study is to examine the most recent research on the validity and reliability of research instruments in mathematics, it is a systematic review, it has used the PRISMA methodology and 15 scientific articles have been reviewed as a sample. This review seeks to strengthen research in the area and to verify the validity and reliability criteria of the instruments. The results are discussed in relation to the comparison of the instruments and their effectiveness in the field of mathematics, in order to provide useful insights for future research and to improve educational practices in this field.

**Keywords** - Validity, reliability, instrument, systematic review.

DOI: 10.18421/TEM134-65

<https://doi.org/10.18421/TEM134-65>

**Corresponding author:** Oscar López- Regalado,  
Graduate School, Cesar Vallejo University, Chiclayo,  
Peru


**Email:** [olopez@ucv.edu.pe](mailto:olopez@ucv.edu.pe)

Received: 03 March 2024.

Revised: 03 May 2024.

Accepted: 02 September 2024.

Published: 27 November 2024

 © 2024 Fani Esperanza Mejía-Clavo, Oscar López- Regalado, Jessica del Rocío Fernández-Velásquez, Antony Esmít Franco Fernández Altamirano, Hilda Aurora Torres Cueva; published by UIKTEN. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

The article is published with Open Access at <https://www.temjournal.com/>

## 1. Introduction

Mathematical skills are essential in education and are valued from childhood in many countries, where their teaching is emphasized through games to stimulate logical thinking from an early age. However, in today's society, mathematical thinking is often perceived as complex and tedious, which can affect child development and lead to imbalances in problem-solving from the preschool stage. It is essential to promote a positive attitude towards mathematics from childhood, using educational methods that encourage active participation and stimulate children's cognitive and logical development. This will ensure a solid foundation for academic success and ability to address challenges in educational and personal future.

In the initial stage, students exhibit a wide range of mathematical skills, encompassing numbers, quantitative thinking, geometry, and problem-solving [1]. Teaching mathematics from an early age is considered crucial to a student's future success by boosting their memory and concentration [2]. At that age, children show math skills through play [3], although they may face difficulties in basic arithmetic skills such as counting and numerical representation in kindergarten. Therefore, it is imperative to teach mathematics in a specialized and didactic way, recognizing its fundamental role in social and motivational development, key aspects to achieve abstract-logical competencies [4]. The development of mathematical learning is intrinsically linked to children's social interaction, and this logical process can be enriched by creative activities that promote their competence evolution [5].

Children's mathematical competence is closely linked to the innovative strategies used by teachers according to the needs of each group of students.

Theories such as Lindmeier's argue that mathematical knowledge begins to develop from the first years of life, accumulating as children explore their environment and understand concepts such as quantity, numerical representations, geometric skills, and shape recognition [6]. These learnings can take place through play, which is an integral part of children's lives [7]. Therefore, it is crucial to employ psychometrically reliable assessment tools during early childhood to assess mathematical skills, as this lays the foundation for formal mathematics education in later years. In addition, the role of executive functions in children's cognitive and academic development is critical, especially in the adoption and retention of new problem-solving strategies in the educational setting [8], [9], [10].

Nowadays, it is necessary to modify various tests to improve mathematical skills, which represent a difficulty for basic education students. Teachers play a crucial role in this regard, promoting and strengthening math connections, [11], [12] that mathematics requires a joint effort of teachers, governments and society in general. Success in learning mathematics is linked to the development of the ability to solve problems and apply skills in everyday situations, which involves making critical judgments, reasoning appropriately, and communicating effectively. It is essential that the tests developed possess validity and reliability, where validity refers to the usefulness of the results obtained and reliability relates to the accuracy of the measurements made [13]. As a result, instruments with validity and reliability in mathematical competencies make it possible to measure dimensions and levels of the area in observable areas, providing information on groups or populations. This article focuses on promoting the search for accuracy in the process of evaluating and communicating messages through the evaluation of results. The main objective is to analyze the research on the validity and reliability of instruments in mathematics in scientific articles from 2019 to 2023, contributing with updated theoretical foundations. The methodology used includes observation and data collection. The central question is: What is the state of research on validity and reliability of research instruments in mathematics in scientific articles from 2019 to 2023? The information was collected from reliable and certified databases to generate an analysis table of relevant articles in the area of mathematics.

## 2. Method

A systematic review and bibliometric analysis was carried out using the PRISMA diagram, which details the methodology and allows comparisons between scientific articles. The search was limited to the years 2019-2023 in six databases, selecting 15 documents in total. For SCOPUS, specific search terms such as "validity" OR "reliability" AND "instrument" OR "questionnaire" OR "ability to solve quantity problems" AND "mathematics" were used, employing Boolean operators for a rigorous search. In Science Direct, BASE, Scielo, Dimensions, and the ERIC database, similar terms focused on the validity of instruments for the area of mathematics were used. It was confirmed that all selected articles were peer-reviewed. This original and up-to-date approach allowed for a comprehensive collection of quantitative and qualitative data relevant to the study.

### 2.1. Inclusion and Exclusion Criteria

Strict criteria were applied for the inclusion of articles in the final sample: a) studies addressing populations ranging from children to adolescents were considered, b) original scientific research articles were prioritized, c) the search was limited to a period of 5 years, d) papers focused on the validation and reliability of instruments in the field of mathematics were selected, e) Data were required to be extracted from the educational context. The exclusion criteria were: a) research presented in book chapters or conferences, b) studies focused on teachers or adults, c) systematic reviews that did not contribute to the topic of interest, d) documents with restricted access, e) works published before 2019. These decisions were made to ensure the consistency and relevance of the selected sample.

### 2.2. Data Collection

From the various databases consulted, 384 relevant articles were initially identified: Scopus (52), Science Direct (26), BASE (91), SciELO (11), ERIC (85) and Dimensions AI (119). After applying filters according to the year of publication, restricting it from 2019, the sample was reduced to 191 articles. After removing duplicates and considering cross-indexing, the sample was reduced to 95 articles.

Subsequently, the researchers reviewed the abstracts of each article, excluding those whose content did not address the constructs of validity and reliability in the area of mathematics, resulting in 47 documents; of these, 13 were excluded for not submitting student-related data, of which 5 came from Scopus and 8 from BASE; Likewise, 6 articles from systematic review were excluded, 9 because they do not contribute to the topic of interest and 4 because of their theoretical nature. There was 100% agreement among the researchers on the exclusion criteria, leaving a final sample of 15 relevant articles (Figure 1).

### 2.3. Year and Country of Studies

In the last 5 years, there has been a level of studies on the validity and reliability of research instruments in mathematics that could be described as limited. It is notable that eight publications were recorded in 2020, while in 2022 there were five and in 2021 only two. These studies come from a total of 11 countries, with individual research conducted in each country, covering 53.28% of the total. Chile and Panama contribute two papers each, representing 26.66% of the total, while Indonesia has three papers, corresponding to 20% of the total. (Figure 2).

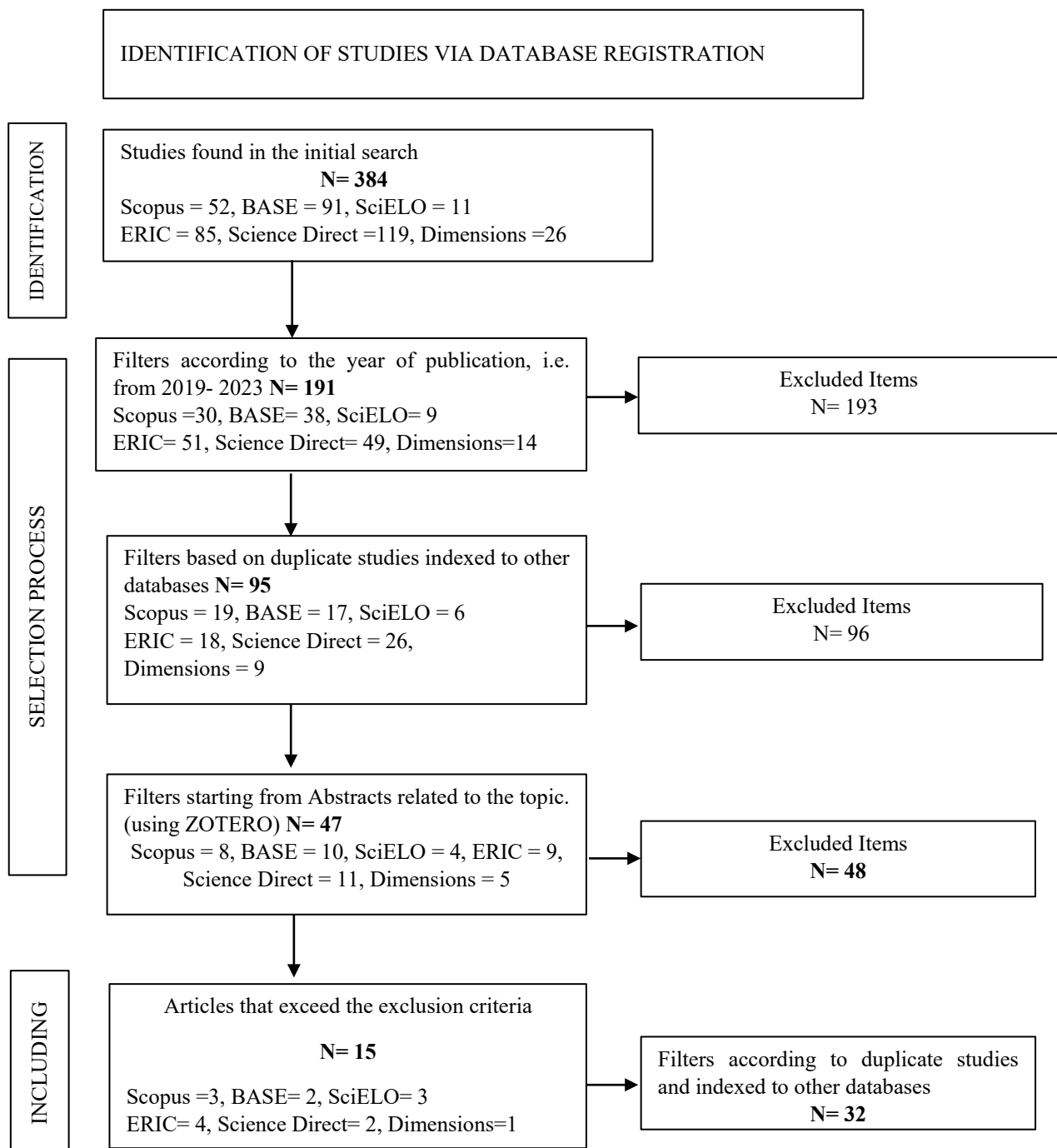


Figure 1. Data collection process in this systematic review

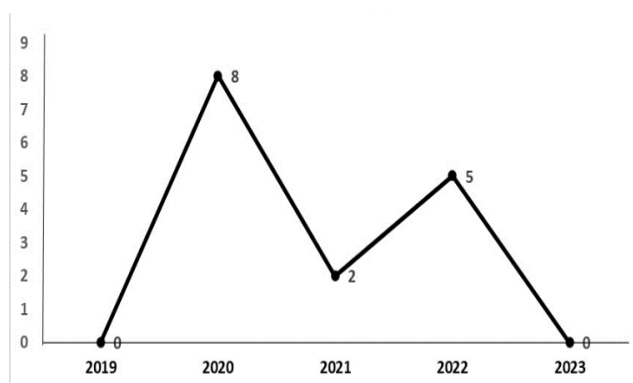


Figure 2. Number of articles per year in the studies carried out

#### 2.4. Sample Size of the Object of Study

The research included in the analysis presented varied samples for the validation and reliability of the instruments, with populations that did not exceed 1200 participants. Specifically, it was found that in 20% of the cases, the samples were less than 150 students, while in 33.3% of the studies, samples of between 150 and 300 participants were used. In 33.3% of the articles, the samples ranged between 300 and 450 students, and in 6.6% of the research, the sample was between 450 and 600 participants. Only 13.33% of the papers exceeded the number of 1000 participants (Figure 3).

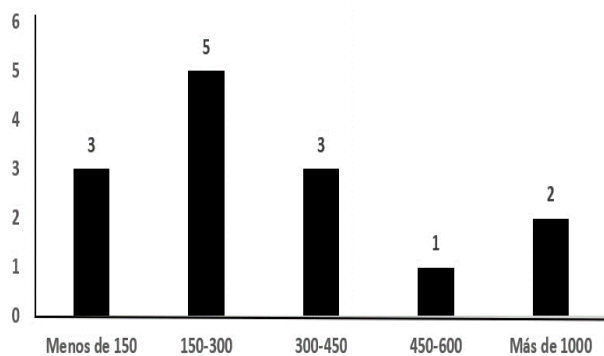


Figure 3. Sample size of the studies reviewed

#### 2.5. Age Characteristics and Other Conditions of the Study Sample

The works reviewed cover all three levels of basic education. Of the 15 articles analyzed, 20% focus on the infant stage, while 46.6% focus on students with equivalent primary schooling. In addition, 20% of the studies include both initial and primary education. Finally, 13.3% of the jobs are related to the secondary level.

#### 2.6. Main Results of the Studies Reviewed

This study reviewed studies from 2019 to July 2023 on the validity and reliability of instruments in mathematics, with the aim of evaluating their impact and progress. The results reveal a considerable scientific production, although there is a lack of specific focus on instrument validation. The 15 papers reviewed come from 11 countries, with diverse samples: less than 150 (3 studies), between 150 and 300 (5), between 300 and 450 (3), between 400 and 600 (1), and more than 1000 (2). The three levels of education (initial, primary, and secondary) are covered. Each article fulfills the purpose of validating and guaranteeing the reliability of instruments in the area of mathematics.

#### 2.7. Coding Data

Various categories of analysis were recorded to obtain the results, including: a) year of publication, b) country of origin of the sample, c) sample size, d) educational stage of the students, e) name of the instrument/items, f) type of validity, g) dimensions and reliability, and h) main results and contributions of the studies.

### 3. Results

The following results were obtained on the 15 articles included in the final sample of the systematic review, in line with the main objective of the study, which consisted of analyzing the research on the validity and reliability of instruments in the area of mathematics (Tables 1 and 2).

Table 1. Methodology of research instruments in the area of mathematics (2019-2023)

Author/ Year	Sample	Stage Educational	Instrument Name/ Items	Type validity/reliability	of Dimensions
(Shavitt <i>et al.</i> , 2022)	565 children	Preschool (4 to 6 years)	Psychometric Tests IDEAL:  Items: 22	Validity by content Cronbach's alpha 0.78.  Reliability (0,961)	Pre-mathematics; Pre-writing; Social-Emotional Development; Self-regulation; Executive Functioning.
(Diaz, 2022)	72 3rd year of secondary school	high school	Test for Types of Math Problems Items = 50	Content Validity Reliability presented a Cronbach alpha of 0.79	Realistic Purely mathematical Fantastic non-routine
(Son, 2022)	50 students	Primary	Model CORE RME	Construct Validity Cronbach's alpha of 0.88 Reliability: Cronbach's alpha of 0.88	Understanding of equivalent representation. Understanding the Relationship of Mathematical Procedures Use of links between mathematical topics. Use math in everyday life.
(Sánchez <i>et al.</i> , 2022)	318 students	Primary (1st to 6th)	Scale to Assess the Reversibility of Thinking in Arithmetic Problems Items = 180	Validity: Confirmatory Factor Analysis Reliability: Cronbach's alpha greater than 70	Change Combination Compare Matching
(Aunio <i>et al.</i> , 2022)	1,029 children	Preschool (3 to 5 years)	Early Arithmetic Test	Validity by content	Observatory Guide
(Canet <i>et al.</i> , 2021)	269 children (9 to 12 years)	Primary	CUFE Executive Functions Questionnaire Items = 32	Internal Validity, Alpha Cronbach's = .79 The total scale had A high level of reliability = .91.	Memory of work Inhibition Cognitive flexibility
(López, <i>et al.</i> , 2021)	366 students	Primary 1st Grade	Early Numeracy Screenner Items = 50	Criterion validity, Cronbach = .903	Counting skills, Numeracy Skills Basic arithmetic skills.
(Gutierrez <i>et al.</i> , 2020)	130 3rd grade students	Primary	Psychometric Tests  Items =23	Validity, analysis, Cronbach's 0.854 reliability, alpha=	Student Learning Readiness Teaching for Student Learning Participation in the management of the school articulated to the community Development of Teaching Professionalism and Identity
(Assel <i>et al.</i> , 2020)	383 children between 4	Preschool And the	CIRCLE Progress	Validity 0.55 to 0.65) Internal consistency	Arithmetic Spatial Ability

	and 9 years old	first grades of elementary school	Monitoring [CPM])		reliability ( $\alpha = 0.94$ ), Test-retest reliability ( $r = 0.78$ )	Geometric Concepts
(Cáceres <i>et al.</i> , 2020)	289 students from 3 to 8 years old	II CYCLE Infants I BASIC CYCLE	EGMA Initial Numeracy Skills Assessment Items =72		Internal Consistency: 0.85 Reliability: 0.48 to 0.86	Number Identification Numerical discrimination. Missing Number Addition Subtraction Problem Solving
(Ashao <i>et al.</i> , 2020)	Elementary school students.	1st grade of primary school	Instrument to measure numeracy skills in real problems at the level of first grade of primary school Items= 28		Content Validity Cronbach's alpha 0.88, satisfactory	Thinking and reasoning. Argumentation Communication Model Building Formulation and problem solving. Use of operations and symbolic, formal and technical language. Representation Use of support and tools
(Hellstrand <i>et al.</i> , 2020)	1139 children between 5 and 8 years old	Garden First and second grade.	Test EN to identify children at risk for math learning difficulties		Validity: Confirmatory Factor Analysis Cronbach's alpha ranged from .70 to .86.	Knowledge of symbolic and non-symbolic numbers Understanding Mathematical Relationships Counting Skills Basic arithmetic skills.
(Setiawan, <i>et al.</i> , 2020)	200 children	Early childhood	creative IM-based curriculum Items = 31.		Validity by construct 0,78	Linguistics Mathematics Visuospatial Kinesthetic Music Interpersonal Intrapersonal Naturalist Religious
(Ashao, <i>et al.</i> , 2020)	159 students	Third Grade Primary	Measurement scale for the identification of geometric skills and competencies  The instrument consists of 24 Items		Content Validity Internal Consistency (0,75)	thought and reasoning, Model building. Problem solving and problem-solving. Argumentation Communication representation Use of operations and formal and symbolic technical language
(Suciati, 2020)	273 students	High school	Math Literacy Tools Items= 10		Validity: Confirmatory Factor Analysis Reliability = 0.78	Observatory guide

Table 2. Main results and contributions of the studies

Author/ Year	Results	Discussion	Conclusion
(Shavitt <i>et al.</i> , 2022)	First, a global analysis of items was performed and correlations between 0.40 and 0.69 points were observed. However, 7 items (items 6, 9, 10, 25, 36, 40, 44) had low correlations (0.19 to 0.39 points).	Executive function assessment questionnaires provide descriptive benchmarks and evidence of their reliability and effectiveness. The final version of the 33-item questionnaire demonstrated reasonable psychometric properties and allowed us to separate the three factors. that explained the variance of 49.08%. 13 items were excluded during the data analysis process.	The conclusion of this study demonstrated that IDELA has reasonable content validity and high inter-rater reliability. Good internal consistency scores were also achieved.
(Diaz, 2022)	As a result, the final version of the test, consisting of 5 problems and 50 related questions, confirms that it is a reliable and valid tool for assessing reading comprehension in the teaching of mathematical quadratic functions.	It should be noted that this study presents a research opportunity under construction and a body of research. Among other possibilities, the effectiveness of this type of tool for the generation of statistics should be highlighted.	An evaluation of the psychometric characteristics of the mathematics test confirms that there is evidence of the adequacy of its content and structure, good identification of its items, and good reliability of the final version of the test.
(Son, 2022)	The CORE RME instrument has acceptable validity and can help observe students' mathematical connections	It was obtained that in the test they showed homogeneous group data on the performance of the students and the improvement in mathematical connectivity.	The conclusion drawn from this study is that the improvement in student performance and mathematical coordination skills through learning the CORE-RME model was greater for students who studied with the CORE model than those who learned with the CORE model. in the traditional model.
(Sánchez <i>et al.</i> , 2022)	An overall reliability analysis was performed to determine data consistency.	Exploratory factor analysis revealed three factors that accounted for more than 70%. A confirmatory factor analysis was performed to analyze the effectiveness of the media., whose index indicated model fit.	The developed instrument is robust enough to evaluate the reversibility of addition and subtraction of elementary operations.
(Aunio <i>et al.</i> , 2022)	The results of reliability and validity justify the use of this test as an early childhood screening test and numerology studies.	The advantage of this test is that it allows us to measure, in a systematic way, the numerical skills of young children that were not possible before. The effects of gender, parents' professional education, residency, and number of siblings are discussed.	Suggestions should be made to direct educational intervention efforts to strengthen teachers' educational practices and increase collegial learning outcomes.

(Canet <i>et al.</i> , 2021)	Consider the evidence of the adequacy of the CUFÉ standards as a safe and valid measure for simple applications to judge children.	Executive function assessment questionnaires provide descriptive benchmarks and evidence of your C and V. The final version of the 33-item questionnaire showed reasonable psychometric properties and allowed us to isolate three main factors that explained the variance of 49.08%. 13 items were excluded during the data analysis process.	Although more research is needed, this study represents a first step toward assessing the executive functions and areas of daily life of school-age children.
(López, <i>et al.</i> , 2021)	Confirmatory factor analysis provided evidence for the three-factor model and confirmed the validity of the constructs	Studies show that math performance is associated with certain aspects (such as counting skills) such as number line measurements, counting, knowledge of numbers, and the ability to add and subtract.	The Early Numeracy Screener serves as an indicator of your toddler's performance in the early levels of math.
(Gutierrez <i>et al.</i> , 2020)	The dimensions of student preparation and student participation were located at a starting level of 33.3% and 40%, respectively)	The analysis revealed differences in the mean values obtained according to the statistical treatment, showing a Pearson correlation of $r = 0.873$ . Being the relationship of the positive variables.	It should be noted that the teacher could use the intervention to relate the math to real problems.
(Assel <i>et al.</i> , 2020)	Exploratory and confirmatory factor analyses revealed that the subtest fits a well-defined model	Concurrent assessments of predictive validity revealed that the CPM math subtest burden correlated with the other assessment loads at a high level ( $r_s 0.55-0.65$ ).	Early recognition of children at risk for math difficulties is important for assessing areas that may be critical for later math development.
(Cáceres <i>et al.</i> , 2020)	The reliability of the EGMA test was calculated considering all study participants. Therefore, the Cronbach value for 96 test items is 0.85, indicating high internal consistency of the instrument.	It is important for teachers to know what strategies to use to help children solve this difficulty in the classroom.	This device is believed to be useful for assessing the mathematics performance of students in the second and first basic cycles of early childhood education.
(Ashao <i>et al.</i> , 2020)	The study yielded an approved device with an index of 0.88 and was considered satisfactory by experts. Also, for the first time in Panama, an instrument was introduced to assess eight cognitive skills in first-grade mathematics.	Olympic-level math problems in primary education demonstrate the need to prepare and train students to develop concrete math exercises and what tools to use.	In line with the theory of learning, an instrument was developed that could measure the content of the ministry's curriculum.
(Hellstrand <i>et al.</i> , 2020)	The test is based on a theoretical model of basic numeracy skills for learning mathematics for children aged 5 to 8 and includes knowledge of symbolic and non-symbolic numbers, understanding of mathematical relationships	The NE test has been found to have adequate construct validity, known group validity, cross-cultural validity, and internal consistency. Empirical data support the purpose of measuring the four numerical skills	The NE test can be considered a sufficient evaluation to identify children at risk for learning problems in mathematics.



(Setiawan, <i>et al.</i> , 2020)	Based on Howard Gardner's theory of multiple intelligences, a creative curriculum assessment model was developed that provides multiple tools to assess the abilities and abilities of infants. This test met the criteria for validity and reliability.	It is recommended that socialization and assessment guides be developed for MI-based creative curriculum learning applications so that the developed products can be used in all early childhood institutions.	The assessment model was useful for early childhood learning.
(Ashao, <i>et al.</i> , 2020)	This result demonstrates the adequacy and validity of the content analysis performed by the panel of experts. An internal consistency analysis resulted in an overall reliability of 0.75, indicating that the designed device has high reliability.	This statistical study has led to the development of a reliable and validated device that can be used in third grade classrooms. Expert opinion indicates that this tool can assess geometric skills and abilities.	The tool consists of 24 items divided into 7 competencies and 4 work situations to ensure that the objectives proposed in the study are achieved. With this tool you will be able to establish new techniques, strategies, teaching and learning methods that promote your students' innate competencies and abilities in real-life situations.
(Suciati, 2020)	This test provides reliable evidence of the validity of structure and content for use as an assessment of learning. Eight experts participated in the study as instrument verifiers	The standardized load factor was greater than the critical limit ( $> 0.3$ ) and the construction reliability (CR) was 0.78 ( $> 0.7$ ).	This test features a high reliability and level of consistency of a good score.

#### 4. Discussion

This work is based on a sample of 15 articles from 11 countries. In the case of Peru, no studies have been found after 2020 on the validity and reliability of instruments in mathematics, with the sole exception of the research of [14]. This limitation in national scientific research restricts the teacher's ability to obtain evidence on performance in the area of mathematics. Since math skills develop cumulatively, with early skills serving as the foundation for later ones, early childhood is critical for development [15]. Therefore, the process of learning mathematics is intrinsically linked to the interaction between students and their social environment. Students naturally monitor their experience and progress in learning mathematics, including the ideas and skills they understand. Problems based on everyday experience play a crucial role in the development of quantitative thinking [16]. The lack of scientific information could be hindering the early detection of problems in the area of mathematics.

In terms of methodological, executive function assessment questionnaires, and the IDELA test have demonstrated their reliability and validity [9], [17]. Significant differences have been identified between subdomain scores, as illustrated in a study with a representative sample of 1029 Finnish children, thus supporting the suitability of these tests for early detection and the study of number sense [18].

On the other hand, study [14] found statistically significant differences in the mean values determined, with a highly positive Pearson correlation [15]. Several studies have linked children's math performance to skills specific to their early numeracy competencies, such as counting, demonstrated in a set of 50 items related to early arithmetic (Cronbach's  $\alpha = .903$ , from our sample). The results of the homogeneity test have supported the consistency of the group's data in terms of achievement and improvement in students' math connections [19]. These findings have allowed the development of a reliable and valid device for implementation in third-grade classrooms, according to the evaluation of experts, who highlight its effectiveness in the assessment of geometric skills.

In the studies reviewed, it is observed that 10 of them have a sample ranging from more than 100 to less than 700 members, while two studies have a sample of less than 60 participants. Finally, two of the 15 tests have a sample of more than 1000 students. Regarding the methodology used, most of the studies (7) adopt a cross-sectional descriptive approach, followed by three descriptive correlational studies. In addition, there is a comparative study with two samples, two standardized studies and two studies belonging to a longitudinal study. In the search carried out, 15 instruments were identified, of which 5 meet the criterion of validity by content, according to studies of [17], [10], [18], [20], [21].

Three of the research papers are construct-valid, specifically the [19], [22]. On the other hand, four of the studies reviewed are valid by confirmatory factor analysis, including the studies of [23], [14], [24]. The remaining three instruments coincide in having internal validity [25], [26], [27]. Regarding the determination of the reliability of the instruments, the 15 studies show a homogeneous reliability, reflected in a Cronbach's alpha coefficient ranging from 0.77 to 0.94. The limitations of the current systematic review focus on the lack of homogeneity in the samples, which cover three levels of basic education. This diversity makes it difficult to draw conclusions about the problem addressed, which highlights the need to validate instruments in this fundamental area. Nonetheless, the information provided by this systematic review offers valuable knowledge to address new scientific challenges in the field.

## 5. Conclusion

The results indicate that, although research on the subject continues, there is a scarcity of works focused on the validation of instruments in the area of mathematics. At the methodological level, there is a variety in the samples used in these studies, which cover different educational levels. Cross-sectional, standardized, quantitative, correlational, comparative, and longitudinal research was carried out with heterogeneous sample sizes. The evaluated instruments show validity by content, by construct, by confirmatory factor analysis, and internal validity, presenting a high reliability for their application.

This research has sought to understand and relate mathematical competence to other areas of knowledge. This review concludes that all the instruments reviewed have dimensions associated with the evaluation of the instrument, although the need for a greater incidence in this type of study is highlighted.

This step represents the continued need to research and develop measurement instruments in the area of mathematics, for the benefit of basic education students. In addition, it underlines the importance of teachers relating mathematics to real problems from an early age, which will help to avoid possible numerical difficulties in the future.

## Acknowledgments

The authors thank the César Vallejo University for funding this research, especially the area of the vice-rector's office for research.

## References:

- [1]. MacDonald, A., & Murphy, S. (2021). Mathematics education for children under four years of age: A systematic review of the literature. *Early Years: An International Journal of Research and Development*, 41, 522-539. Doi: 10.1080/09575146.2019.1624507
- [2]. Balt, M., Fritz, A., & Ehlert, A. (2020, June). Insights into first grade Students' development of conceptual numerical understanding as drawn from progression-based assessments. In *Frontiers in Education*, 5, 80. Frontiers Media SA. Doi:10.3389/educ.2020.00080
- [3]. Reikerås, E. (2020). Relations between play skills and mathematical skills in toddlers. *ZDM*, 52(4), 703-716. Doi: 10.1007/s11858-020-01141-1
- [4]. Wästerlid, C. (2022). Low-achieving grade K-3 children's early numeracy competences: A systematic literature review. *International Journal of Early Years Education*, 30(4), 748-765. Doi: 10.1080/09669760.2020.1848524
- [5]. Nur, A. S., Kartono, K., Zaenuri, Z., & Rochmad, R. (2022). The learning trajectory construction of elementary school students in solving integer word problems. *Participatory Educational Research*, 9(1), 404-424.
- [6]. Lindmeier, A., Seemann, S., Kuratli-Geeler, S., Wullschleger, A., Dunekacke, S., Leuchter, M., Vogt, F., Opitz, E. M., & Heinze, A. (2020). Modelling early childhood teachers' mathematics-specific professional competence and its differential growth through professional development – an aspect of structural validity. *Research in Mathematics Education*, 22(2), 168-187. Doi: 10.1080/14794802.2019.1710558
- [7]. Rezeki, S., Andrian, D., & Safitri, Y. (2021). Mathematics and cultures: A new concept in maintaining cultures through the development of learning devices. *International Journal of Instruction*, 14(3), 375-392. Scopus. Doi: 10.29333/iji.2021.14322a
- [8]. Cáceres, S, Sepúlveda, F., & Rodríguez, C. (2020). Standardization of the EGMA Instrument for the Assessment of Initial Numerical Skills in Basic Education Students in Chile. *Estudios pedagógicos (Valdivia)*, 46(1), 301-318. Doi: 10.4067/S0718-07052020000100301
- [9]. Canet-Juric, L., Verónica del-Valle, M., Gelpi-Trudo, R., García-Coni, A., Vanesa Zamora, E., Introzzi, I., & Laura Andrés, M. (2021). Development and Validation of the Executive Function Questionnaire (cufe) for Children aged 9-12 Years. *Avances en Psicología Latinoamericana*, 39(1). Doi: 10.12804/revistas.urosario.edu.co/apl/a.9892

- [10]. Díaz, V. (2022). Design and Validation of a Test for the Types of Mathematical Problems Associated with Reading Comprehension. *Education Sciences*, 12(11). Scopus. Doi: 10.3390/educsci12110795
- [11]. Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM*, 52(1), 1-16. Doi: 10.1007/s11858-020-01130-4
- [12]. Moral, L., García, E., Duran, R. (2019). Formative intervention for mathematics learning: An approach from a diploma course. *Conrad*, 15(69), 7-18.
- [13]. Medina, J., Ramírez, M., & Miranda, I. (2019). Validity and reliability of an online test. *Apertura (Guadalajara, Jal.)*, 11(2), 104-121. Doi: 10.32870/ap.v11n2.1622
- [14]. Gutierrez, A. B., Palacios, J., Fuster Guillen, D., & Hernandez, R. (2020). Evaluation of teaching performance in the learning achievement of the mathematics area. *Revista De Psicología*, 16(31), 48-57.
- [15]. Lopez-Pedersen, A., Mononen, R., Korhonen, J., Aunio, P., & Melby-Lervåg, M. (2021). Validation of an Early Numeracy Screener for First Graders. *Scandinavian Journal of Educational Research*, 65(3), 404-424. Doi: 10.1080/00313831.2019.1705901
- [16]. Saparuddin, A., Kartono, K., Zaenuri, Z., & Rochmad, R. (2022). The Learning Trajectory Construction of Elementary School Students in Solving Integer Word Problems. *Participatory Educational Research*, 9(1), 404-424. Doi: 10.17275/per.22.22.9.1
- [17]. Shavitt, I., de Araujo Scatollin, M. A., Rossi, A. S. U., Mercadante, M. P., Gamez, L., Resegue, R. M., ... & do Rosário, M. C. (2022). Transcultural adaptation and psychometric properties of the International Development and Early Learning Assessment (IDELA) in Brazilian pre-school children. *International Journal of Educational Research Open*, 3, 100138. Doi: 10.1016/J.Ezeron.2022.100138
- [18]. Aunio, P., Hautamäki, J., Heiskari, P., & Van Luit, J. E. (2006). The early numeracy test in finnish: Children's norms. *Scandinavian Journal of Psychology*, 47(5), 369-378.
- [19]. Son, A. L. (2022). The Students' Abilities on Mathematical Connections: A Comparative Study Based on Learning Models Intervention. *Mathematics Teaching Research Journal*, 14(2), 72-87.
- [20]. Ashaw, M., Gutiérrez, J., Tuñón, R., Bethancourt, B., Beitía, J., & Cruz, N. (2020). Validation of an instrument to measure numeracy skills in real problems at the level of the first grade of primary school. *Conrad*, 16(73), 91-99.
- [21]. Ashaw, M., Gutiérrez, J., Zamora, E., & Cesar, Y. (2020). Escala de medición para la identificación de habilidades y competencias geométricas. Enfoque de problemas en contexto real en el tercer grado primaria. *Revista Conrad*, 16(77), 490-494.
- [22]. Setiawan, R., Mardapi, D., & Karyanto, U. B. (2020). Multiple Intelligences-Based Creative Curriculum: The Best Practice. *European Journal of Educational Research*, 9(2), 611-627.
- [23]. Sánchez-Barbero, B., Cáceres, M. J., & Chamoso, J. M. (2022). Validation of a Scale to Assess the Reversibility of Thought in Verbal Arithmetic Problems. *Acta Scientiae*, 24(6), 370-401. Doi: 10.17648/acta.scientiae.7328
- [24]. Hellstrand, H., Korhonen, J., Räsänen, P., Linnanmäki, K., & Aunio, P. (2020). Reliability and validity evidence of the early numeracy test for identifying children at risk for mathematical learning difficulties. *International Journal of Educational Research*, 102, 101580. Doi: 10.1016/j.ijer.2020.101580
- [25]. Canet, J., Del Valle, M., Gelpi, T., García, A., Zamora, E., Introzzi, I., & Andrés, M. (2021). Development and validation of the Executive Function Questionnaire in children aged 9 to 12 years. *Advances in Latin American Psychology*, 39(1). Doi: 10.12804/revistas.urosario.edu.co/apl/a.9892
- [26]. Assel, M. A., Montroy, J. J., Williams, J. M., Foster, M., Landry, S. H., Zucker, T., Crawford, A., Hyatt, H., & Bhavsar, V. (2020). Initial Validation of a Math Progress Monitoring Measure for Prekindergarten Students. *Journal of Psychoeducational Assessment*, 38(8), 1014-1032. Doi: 10.1177/0734282920922078
- [27]. Suciati, Munadi, S., Sugiman, & Ratna Febriyanti, W. D. (2020). Design and validation of mathematical literacy instruments for assessment for learning in Indonesia. *European Journal of Educational Research*, 9(2), 865-875. Scopus.