LEARNING : Jurnal Inovasi Penelitian Pendidikan dan Pembelajaran
Vol. 4 No. 4 November-Desember 2024
E-ISSN : 2777-0575
P-ISSN : 2777-0583
Online Journal System : https://jurnalp4i.com/index.php/learning

PSYCHOMETRIC ANALYSIS OF MATHEMATICS READINESS ASSESSMENT TOOL FOR ELEMENTARY SCHOOL STUDENTS: A RASCH MODEL APPROACH

RIKY SHEPTIAN*, RIYAN ARTHUR, RAHAYU WARDHANI
Jakarta State University, Jakarta
e-mail: ricky01abiogenesis@gmail.com

ABSTRAK

This study examines the psychometric properties of a Mathematics Readiness Assessment Tool for elementary school students using Rasch analysis through the Winstep program. The analysis involved 214 students responding to 15 items. Results show strong psychometric characteristics with an item reliability of 0.95 and person reliability of 0.69. The instrument demonstrates good construct validity with INFIT and OUTFIT MNSQ values falling within the acceptable range (0.5-1.5). Analysis of unidimensionality showed satisfactory results with raw variance explained by measures at 26.8% and unexplained variance in the first contrast at 7.6%. Local independence assumption was met with residual correlations ranging from -0.22 to 0.20. Item difficulty ranges from -1.33 to 1.56 logits, indicating a good spread of item difficulties. The test information function shows optimal measurement precision for students of average ability levels, though less precise for extremely high or low abilities. Differential Item Functioning (DIF) analysis revealed some items requiring revision, particularly items 13-15. The instrument's Alpha Cronbach value of 0.73 indicates good internal consistency. Overall, the assessment tool demonstrates adequate psychometric properties for measuring mathematics readiness in elementary school students, though some improvements are recommended for optimal measurement across all ability levels.

Keywords: *Mathematics readiness, Rasch analysis, psychometric properties, elementary school, assessment tool*

INTRODUCTION

Mathematics education at the elementary school level plays an important role in building the foundation of students' numeracy and problem-solving skills. These abilities are not only prerequisites for understanding more complex mathematical concepts at the next level of education, but also play a role in developing critical and logical thinking skills in children (National Council of Teachers of Mathematics, 2000). Therefore, mathematics readiness assessment is very necessary to understand the extent to which students are ready to learn mathematics, both in terms of basic abilities and initial understanding.

In the context of educational assessment, accurate and reliable measurement is the main key to getting a real picture of student abilities. One approach that can be used in analyzing assessment instruments is the Rasch Model, which is included in modern psychometric analysis. The Rasch Model offers advantages in measuring student abilities and validating instruments based on the level of item difficulty and respondent abilities objectively (Bond & Fox, 2015). This approach allows the development of more precise measuring instruments, in accordance with the principles of educational measurement and provides deeper interpretation.

The use of the Rasch Model in the analysis of mathematics readiness at the elementary school level is also important considering the challenges of heterogeneity in student abilities. By analyzing the assessment instrument through this approach, weaknesses in the test items can be identified, and the instrument can be adjusted to suit the characteristics of the students. Several previous studies have shown that the Rasch Model approach is effective in assessing the validity and reliability of educational assessment instruments and providing more detailed

LEARNING : Jurnal Inovasi Penelitian Pendidikan dan Pembelajaran
Vol. 4 No. 4 November-Desember 2024
E-ISSN : 2777-0575
P-ISSN : 2777-0583
Online Journal System : https://jurnalp4i.com/index.php/learning

information regarding the suitability of test items to students' abilities (Boone, Staver, & Yale, 2014).

The Rasch measurement model offers sophisticated analytical capabilities for evaluating assessment tools, providing detailed information about item characteristics, test functioning, and measurement precision. This analysis examines several critical aspects:

1. Item and person reliability and separation indices
2. Unidimensionality and local independence
3. Item difficulty distribution and targeting
4. Test information function
5. Differential item functioning across groups

Previous research has shown that well-designed mathematics readiness assessments can effectively predict student success in elementary mathematics. However, there remains a need for thoroughly validated instruments that can accurately measure mathematical readiness while meeting rigorous psychometric standards.

This analysis aims to:

- Evaluate the psychometric properties of the Mathematics Readiness Assessment Tool
- Assess the instrument's effectiveness in measuring student mathematical readiness
- Identify potential areas for improvement in the assessment
- Provide evidence-based recommendations for the tool's use in educational settings

The findings from this analysis will contribute to the improvement of mathematics readiness assessment practices and support evidence-based decision-making in early mathematics education.

RESEARCH METHOD

This study employed a quantitative psychometric analysis design using the Rasch measurement model. The analysis focused on evaluating the psychometric properties of a Mathematics Readiness Assessment Tool designed for elementary school students. The study involved 214 students as participants (N=214) who were assessed using the mathematics readiness instrument. The sample size was adequate for Rasch analysis, providing sufficient data points for stable item calibration and person measurement estimates.

The Mathematics Readiness Assessment Tool consisted of 15 items measuring mathematical competencies required for elementary school readiness. The instrument utilized a dichotomous scoring system (0-1), where:

- Score 0: Incorrect response
- Score 1: Correct response

Data was collected through the administration of the Mathematics Readiness Assessment Tool to the participants. The assessment was conducted following standardized procedures to ensure consistency in test administration and data collection. Data analysis was conducted using the Winstep software version 3.73, implementing the Rasch measurement model. The analysis included several key components:

1. **Reliability Analysis**
    - Person reliability
    - Item reliability
    - Separation indices
    - Cronbach's Alpha
2. **Validity Analysis**
    - Construct validity through INFIT and OUTFIT MNSQ statistics
    - Principal Component Analysis of residuals for unidimensionality
    - Local independence assessment

- Item characteristic curve analysis

3. **Item Analysis**
   - Item difficulty estimates
   - Item fit statistics
   - Differential Item Functioning (DIF)
   - Item characteristic curves
4. **Test Function Analysis**
   - Test information function
   - Test characteristic curve
   - Wright map analysis
   - Person-item distribution
5. **Quality Control Criteria**
   - INFIT and OUTFIT MNSQ acceptable range: 0.5 - 1.5
   - ZSTD acceptable range: -2.0 to +2.0
   - Point measure correlation: positive and > 0.3
   - Unidimensionality: unexplained variance in first contrast < 15%
   - Local independence: residual correlations < ±0.70

The analysis process involved iterative procedures with 5 iterations to achieve stable calibration, as indicated by decreasing MAX SCORE RESIDUAL (from -2.90 to -0.15) and MAX LOGIT CHANGE (from 0.0616 to 0.0033).

Data Interpretation

Results were interpreted based on established psychometric criteria and guidelines for Rasch analysis. Special attention was given to:

- Reliability coefficients
- Fit statistics
- Item difficulty distribution
- Test information function
- Evidence of construct validity
- DIF patterns
- Wright map person-item distributions

This comprehensive analysis approach provided detailed information about the instrument's psychometric properties and its effectiveness in measuring mathematics readiness among elementary school students.

RESULTS AND DISCUSSION

The psychometric analysis of the Mathematics Readiness Assessment Tool for elementary school students has yielded findings that can be explained and supported by various theoretical frameworks. These results will be discussed in the context of measurement theory, cognitive development theory, mathematical learning theory, and educational assessment theory.

Reliability and Validity in the Context of Measurement Theory

Table 1 Summary of Reliability and Separation Statistics

Statistic	Person	Item
Reliability	0.69	0.95
Separation	1.49	4.56
INFIT MNSQ	1.00	1.02
OUTFIT MNSQ	0.94	0.94
Cronbach Alpha	0.73	-

The reliability analysis results (Table 1) show excellent item reliability (0.95) and adequate person reliability (0.69). These findings can be explained through Modern Measurement Theory developed by Rasch (1960), which emphasizes the importance of measurement invariance. According to Rasch, good measurement should produce consistent results regardless of the sample used. The high item reliability confirms this principle.

Wright & Stone (1979) further developed this concept by emphasizing that good reliability reflects the stability of item difficulty hierarchy. The Cronbach's Alpha value of 0.73 indicates good internal consistency, aligning with Classical Test Theory developed by Spearman (1904) and extended by Cronbach (1951), which states that reliable measurement must show consistency in measuring the same construct.

Unidimensionality and Theoretical Constructs

Table 2. Summary of Unidimensionality Analysis

Component	Value
Raw variance explained by measures	26.8%
Unexplained variance in 1st contrast	7.6%
Unexplained variance in 2nd contrast	7.5%
Unexplained variance in 3rd contrast	6.8%
Unexplained variance in 4th contrast	6.3%
Unexplained variance in 5th contrast	5.9%

The unidimensionality analysis (Table 2) showing 26.8% variance explained by measurement can be understood through constructivism theory in mathematics learning. The APOS Theory (Action, Process, Object, Schema) developed by Dubinsky & McDonald (2001) explains that mathematical understanding develops through interrelated stages. The unexplained variance in the first contrast of only 7.6% supports the notion that mathematical readiness is a coherent construct.

The RME (Realistic Mathematics Education) Model developed by Freudenthal (1991) emphasizes that mathematics should be connected to reality and represents human activity. The good unidimensionality findings indicate that the instrument successfully measures a meaningful and integrated mathematical construct.

Item Difficulty Levels and Developmental Theory

Table 3. Item Difficulty and Fit Statistics

Item	Measure	Error	INFIT MNSQ	OUTFIT MNSQ
Q15	1.56	0.18	1.11	0.89
Q13	0.95	0.17	0.97	0.88
Q14	0.81	0.16	0.76	0.68
Q6	0.28	0.16	1.13	1.12
Q4	0.18	0.16	0.88	0.86
Q9	0.16	0.16	1.11	1.05
Q10	0.11	0.16	1.11	1.10
Q5,7,11	0.08	0.16	0.92	0.88
Q12	-0.04	0.16	0.77	0.69
Q8	-0.49	0.16	1.12	1.17
Q3	-1.13	0.17	1.08	0.93
Q2	-1.30	0.17	1.14	0.96
Q1	-1.33	0.17	1.08	0.95

The distribution of item difficulties (Table 3) ranging from -1.33 to 1.56 logits reflects the cognitive developmental stages explained in Piaget's theory (1970). Piaget identified the concrete operational stage (7-11 years) as a critical period in the development of mathematical understanding. The item difficulty hierarchy in this instrument aligns with the development from concrete to semi-abstract thinking characteristic of elementary school students.

Van Hiele (1986) developed a specific theory about the development of geometric understanding that can be applied to mathematical understanding in general. Van Hiele's five levels of understanding (visualization, analysis, informal deduction, formal deduction, and rigor) are reflected in the variation of item difficulties, where easy items require simple visualization while difficult items require more complex analysis.

Test Information Function and Modern Measurement Theory

Table 4. Test Information Function Values

Theta Level	Information	Standard Error
-3.0	0.5	1.41
-2.0	1.0	1.00
-1.0	2.0	0.71
0.0	3.3	0.55
1.0	3.0	0.58
2.0	1.5	0.82
3.0	0.7	1.20

The test information function analysis (Table 4) shows optimal measurement at average ability levels (theta 0). This finding aligns with Item Response Theory (IRT) developed by Lord (1980). According to Lord, the test information function reflects measurement precision at various ability levels. The maximum information value of 3.3 at theta 0 with minimum standard error of 0.55 indicates precise measurement for average-ability students.

Embretson & Reise (2000) expanded the understanding of information functions in educational contexts, emphasizing that good tests should provide maximum information at ability levels consistent with measurement objectives. In the context of elementary mathematics readiness, optimal measurement at average ability aligns with the need to identify students requiring additional support.

Mathematical Readiness Theory and Learning Trajectories

Clements & Sarama (2014) developed learning trajectories theory in early mathematics emphasizing that mathematical learning follows predictable developmental paths. The observed item difficulty distribution reflects these trajectories:

- Easy items (Q1-Q3, -1.33 to -1.13 logits) represent foundational skills
- Medium items (Q4-Q12, -0.04 to 0.28 logits) reflect developing understanding
- Difficult items (Q13-Q15, 0.81 to 1.56 logits) represent advanced comprehension

Jordan et al. (2009) identified critical components in early mathematical readiness:

1. Number mastery
2. Quantitative reasoning
3. Calculation skills

The item difficulty structure in this instrument aligns with these components, supporting the construct validity of the instrument.

Differential Item Functioning and Measurement Fairness

Table 5. Summary of DIF Analysis

Item	DIF Size	t-statistic	Probability
Q13	2.5	3.2	0.002
Q14	2.3	3.0	0.003

Q15	2.4	3.1	0.002
Q7	1.0	1.5	0.134
Q8	0.9	1.4	0.162

The DIF analysis (Table 5) showing variation in item function across groups can be understood through Thompson et al.'s (2002) Universal Design for Assessment framework. This theory emphasizes the importance of accessibility and fairness in educational measurement. Significant DIF in items 13-15 indicates the need for revision to meet these principles.

Messick (1989) developed validity theory emphasizing the social consequences of measurement. The DIF findings demonstrate the importance of considering fairness aspects in instrument development, consistent with Messick's perspective on consequential validity.

Integration with Vygotskian Theory and Zone of Proximal Development

Vygotsky's (1978) Zone of Proximal Development (ZPD) concept explains the distance between students' actual and potential developmental levels. The test information function's optimality at average ability supports the instrument's use in identifying students' mathematical ZPD. Heddens (1964) further developed this concept in mathematics learning through concrete-semiconcrete-abstract level theory.

Theoretical Implications for Educational Practice

The integration of these theoretical perspectives yields several important implications:

1. Theory-Based Measurement
- The instrument demonstrates alignment with modern measurement principles
- Reliability and validity support educational decision-making use
- Psychometric structure reflects underlying theoretical constructs

2. Mathematical Development
- Item hierarchy supports staged developmental models
- Results align with constructivist mathematics learning theory
- Optimal measurement at critical levels for educational intervention

3. Fairness and Accessibility
- DIF findings indicate areas for improvement
- Adjustments needed to meet universal design principles
- Importance of considering socio-cultural context in measurement

Theoretical Limitations and Development Directions

While results support various existing theories, several limitations warrant consideration:

1. Moderate person reliability indicates the need for theoretical development regarding individual variability in mathematical readiness
2. Limited precision at extreme abilities suggests the need for theoretical elaboration of high-level mathematical understanding
3. DIF issues indicate the need for further integration between measurement theory and educational equity theory

Further development should consider synthesizing modern measurement theory, cognitive developmental theory, and educational equity principles to produce more comprehensive and fair instruments.

Integration with Current Research

This analysis contributes to contemporary research in:

1. Early mathematics assessment (Dong & Chen, 2020)
2. Psychometric analysis of educational instruments (Boone et al., 2014)
3. Mathematics readiness evaluation (Andrews & Sayers, 2021)

The findings suggest that while the instrument demonstrates strong psychometric properties, ongoing theoretical development is needed to address identified limitations and enhance measurement precision across all ability levels.

Discussion

Based on the comprehensive psychometric analysis of the Mathematics Readiness Assessment Tool for elementary school students using the Rasch model through Winstep software, this study has demonstrated the overall effectiveness and validity of the instrument while identifying specific areas for enhancement. The analysis revealed strong psychometric properties, with an excellent item reliability index of 0.95 and good internal consistency as shown by a Cronbach's Alpha of 0.73. The instrument successfully meets key measurement criteria, including satisfactory fit statistics with INFIT and OUTFIT MNSQ values falling within the acceptable range of 0.5-1.5, and strong evidence of unidimensionality with 26.8% of variance explained by measures.

The assessment tool effectively measures mathematical readiness across a broad ability spectrum, ranging from -3 to +3 logits, with particularly strong measurement precision for students of average ability levels. The item difficulty distribution, spanning from -1.33 to 1.56 logits, provides good coverage of the ability continuum and maintains local independence among items. This distribution allows for effective differentiation of student abilities, particularly in the middle range of the measurement scale, making it well-suited for general screening purposes in elementary school settings.

However, several areas for improvement have been identified through this analysis. The moderate person reliability of 0.69, while acceptable, suggests room for enhancement in the consistency of person measurement. Significant Differential Item Functioning (DIF) observed in items 13-15 indicates a need for revision to ensure fair assessment across different student groups. Additionally, the analysis revealed gaps in the item difficulty distribution, particularly for high-ability students, suggesting the need for additional items to provide more precise measurement at the upper end of the ability spectrum.

From a practical standpoint, the instrument demonstrates strong utility for its intended purpose of assessing mathematics readiness in elementary school students. It is particularly effective for general screening and identifying students who may need additional support in their mathematical development. However, educators should be aware that supplementary assessment tools may be necessary for students at the extremes of the ability range, both very high and very low, where the current instrument shows reduced measurement precision.

These conclusions support the instrument's implementation in educational settings while providing clear direction for future development. Specifically, future revisions should focus on improving person reliability, addressing DIF issues in identified items, and developing additional items to better assess high-ability students. Such enhancements would further strengthen the instrument's effectiveness in supporting educational decision-making and ensuring appropriate mathematical support for all students transitioning into elementary school.

Finally, this study contributes to the broader field of educational measurement by demonstrating the effective application of Rasch analysis in validating mathematics readiness assessments. The findings provide both theoretical insights into measurement properties and practical guidance for instrument development, supporting evidence-based approaches to educational assessment.

CONCLUSSION

The Mathematics Readiness Assessment Tool for elementary school students has been analyzed using the Rasch model and Winstep software. The tool has strong psychometric properties, with an excellent item reliability index of 0.95 and good internal consistency. It measures mathematical readiness across a broad ability spectrum, with strong measurement precision for average ability levels. However, areas for improvement include moderate person reliability of 0.69, significant Differential Item Functioning (DIF) in items 13-15, and gaps in the item difficulty distribution, particularly for high-ability students. The instrument is effective for general screening and identifying students needing additional support in their mathematical development. Future revisions should focus on improving person reliability, addressing DIF issues, and developing additional items to better assess high-ability students. This study contributes to the field of educational measurement by demonstrating the effectiveness of Rasch analysis in validating mathematics readiness assessments.

DAFTAR PUSTAKA

Adams, R. J., & Wu, M. L. (2010). *Multidimensional Models for Measuring Student Progress*. Springer.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer.

Clements, D. H., & Sarama, J. (2014). *Learning and teaching early math: The learning trajectories approach* (2nd ed.). Routledge.

De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. Guilford Press.

Dubinsky, E., & McDonald, M. A. (2001). APOS: A constructivist theory of learning in undergraduate mathematics education research. In D. Holton (Ed.), *The teaching and learning of mathematics at university level* (pp. 275-282). Springer.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*(3), 850-867.

Linacre, J. M. (2012). *A User's Guide to WINSTEPS: Rasch-Model Computer Programs*. Winsteps.com.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. NCTM.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). University of Minnesota, National Center on Educational Outcomes.

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Routledge.