International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 01 January 2026
DOI: 10.47191/ijcsrr/V9-i1-16, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

Unidimensionality Analysis and Differential Item Functioning in the Applied Mathematics Final Examination Test of Politeknik Negeri Bali Students Using the Rasch Model

I Ketut Darma[1*], Adi Herdiansyah[2], Arnaningtyas Rofi'i[3]
[1,2,3]Department of Mechanical Engineering, Politeknik Negeri Bali, Indonesia

ABSTRACT: The objectives of this research are (1) to analyze the fulfillment of unidimensionality assumptions on Applied Mathematics FSE instruments by the Rasch Model, (2) to identify and analyze the existence of items that have differential item functioning (DIF) among study programs in the polytechnic, and (3) to interpret as well as recommend improvements to instruments based on results from analysis about unidimensionality and DIF. This survey is quantitative research with a cross-sectional design. The sample consists of 206 students coming from three study programs: the Mechanical Engineering, the Refrigeration and Air Conditioning Engineering, and the Utility Engineering Technology, Politeknik Negeri Bali (PNB), selected by the total sampling method. The research instrument comprises five questions constructed according to an outcome-based education (OBE) curriculum where content validity was checked through Aiken's V method, giving a value equal to 0.91. Data were analyzed using Winsteps 5.9 software based on the Rasch model. Results of the analysis indicated that the test has outstanding reliability and separation indices both on respondents and items. Most items fall within the fitting criteria of the Rasch Model. The unidimensionality test proved that it can consistently measure a single construct in applied mathematics competency. Based on DIF analysis, most items work well across different learning categories, but a few items have large and significant DIFs. These imply that this instrument is suitable to be used as an assessment tool for learning outcomes where several parts still need improvement to enhance fairness in measuring. It raises awareness about conducting comprehensive evaluations on assessment tools within the OBE context in vocational education.

KEYWORDS: Applied Mathematics, Differential Item Functioning, Rasch Model, Unidimensionality, Vocational Education

INTRODUCTION

Summative assessments, such as the Applied Mathematics Final Exam (FSE) in measuring learning outcomes in vocational higher education, are crucial. Their role is to ensure student competency achievement. So, assessment tools need to be very clear about what they are supposed to measure. Applied mathematics courses form the foundation for several areas of technical competence and problem-solving within engineering disciplines [1]. Therefore, this Applied Mathematics FSE instrument needs to be validated through detailed psychometric analysis that will objectively ensure the actual abilities reflected in students' scores.

The Rasch model is a modern analytical approach used to evaluate instrument quality in greater depth. This model allows researchers to assess item suitability, reliability, difficulty level, unidimensionality, and potential intergroup bias with a higher degree of objectivity than classical approaches [2,3]. This model is also applicable to applied mathematics because it provides a v[4]ry objective and fair way to measure student performance, mapping students and tasks to the same logical scale, so that quality measurement does not depend on the number or difficulty of the tasks [2,3]. This model allows for in-depth item analysis to find items that are too easy, too hard, or not appropriate for the construct being measured. It also makes sure that the test is unidimensional, as basic measurement theory says it should be [5]. Furthermore, Differential item functioning (DIF) testing helps ensure fairness across student groups, ensuring that no group is advantaged or disadvantaged by specific item characteristics [6]. The Rasch model provides the necessary diagnostic information to effectively support vocational education assessments in implementing outcome-based education (OBE) and Continuous Quality Improvement. This model strengthens the validity, reliability, and fairness of applied mathematics exams, ensuring that assessment results are sufficiently accurate to serve as a basis for academic decision-making [4,7,8].

The key assumptions underlying the Rasch Model are unidimensionality and DIF. Unidimensionality means that all items must measure the same ability construct [2]. If this assumption is not met, interpretation of test results can be misleading because the scores

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 01 January 2026
DOI: 10.47191/ijcsrr/V9-i1-16, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

obtained do not reflect a single ability [2,9]. Meanwhile, DIF the fairness aspect of an instrument, assesses the balance and fairness of a test across groups of respondents, identifies potentially biased items, and improves the construct validity and overall fairness of the instrument [10,11]. The presence of DIF implies bias, so the instrument no longer meets the principle of fairness [12]. Polytechnic students consist of various study program groups such as Mechanical Engineering, Refrigeration and Air Conditioning Engineering, and Utility Engineering Technology. DIF analysis can test whether there are items that benefit or disadvantage certain groups even though their abilities are equal. Fair questions for groups of students from Mechanical Engineering, Refrigeration and Air Conditioning Engineering, or Utility Engineering Technology study programs.

However, most previous research on the quality of Applied Mathematics tests in Polytechnics has focused solely on item suitability or reliability analysis, without in-depth testing of unidimensionality and DIF. Furthermore, DIF analysis generally focuses only on gender groups, not on vocational education characteristics such as study program differences. Research on Rasch analysis of Applied Mathematics tests in Indonesian vocational higher education institutions is also scarce, despite the implementation of the OBE curriculum, which requires assurance of validity and fairness of assessment.

This study aims to: 1) analyze the fulfillment of the unidimensionality assumption in the Applied Mathematics FSE test using the Rasch Model; 2) identify and evaluate the presence of items containing DIF between study programs at the Polytechnic; and 3) provide interpretations and recommendations for test improvement based on the results of the unidimensionality and DIF analysis.

This article presents unidimensionality and DIF tests on the Applied Mathematics FSE test used in Polytechnics, making a comparison between the Mechanical Engineering, Refrigeration and Refrigeration Engineering, and Utility Engineering Technology study programs that are rarely observed from the aspect of item bias. Therefore, this study provides new empirical evidence on Rasch PCA residuals and Rasch-based DIF analysis in evaluating the quality of instruments in vocational higher education in Indonesia.

RESEARCH METHODS

This study adopted a quantitative approach using a cross-sectional design because the latter is one of the most common methodologies in educational assessment research that permits simultaneous data collection and analysis at one point in time. It is appropriate in judging the characteristics of assessment instruments and participant performance without instituting any intervention [13,14].

The population in this study were all even-semester students of the 2023/2024 academic year, Department of Mechanical Engineering, PNB who took the Applied Mathematics FSE. Sampling used the total sampling technique, namely, all students who took the exam and provided complete and valid answers. Thus, the total sample was 206 people, namely, 117 students of the Mechanical Engineering Study Program (1), 72 students of Refrigeration and Air Conditioning Engineering (2), and 17 students of Utility Engineering Technology (3).

.

In this study, 40 multiple-choice applied mathematics FSE questions were analyzed. The questions were arranged based on the Semester Learning Plan, which included limits, derivatives, and integrals. The questions were arranged by a team of lecturers who teach the course and were officially used in the implementation of the FSE in the Department of Mechanical Engineering. Before use, this instrument was tested through a content validation process by experts. The validation process used the Aiken's V method. The V coefficient value was 0.91, indicating that the instrument had a very high level of content validity and was suitable for use [15].

Data were collected during the implementation of the even-semester FSE of the 2023/2024 academic year through PNB e-learning, and questions were presented in Google Forms. Although online, the exam was conducted in class under the direct supervision of two supervisors to maintain the integrity of the exam. Subsequently, the results of the students' answers were exported in the form of a digital spreadsheet and coded with a value of 1 for correct answers and 0 for incorrect answers. The data were analyzed using Rasch modeling assisted by Winsteps 5.9 software with a focus on item and person fit, reliability and index separation, unidimensionality, and DIF.

Reliability and Separation Index

The Rasch model uses three components to measure the reliability of an instrument: Cronbach's alpha (KR-20), personnel and item reliability, and personnel and item separation. This study used item and respondent level reliabilities, including the calculation of

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 01 January 2026
DOI: 10.47191/ijcsrr/V9-i1-16, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

the separation index, to show how well an instrument can discriminate levels of competence or item difficulty. The person and item reliability values may be interpreted as follows: very high for above 0.90, high from 0.80 to 0.89, adequate from 0.70 to 0.79, low from 0.60 to 0.69, very low (poor) from 0.50 to 0.59, and unacceptable below 0.50. The classification of person and item separation values for the value range: ≥ 3.00 is classified as very high, 2.00–2.99 as high, 1.00–1.99 as sufficient, < 1.00 as low [9,16].

Item and Respondent Suitability

To ensure that the project fully measures the constructs, we evaluated the project fit of the output project using three main metrics of the Rasch model: mean square fit (MNSQ), Z-standard fit (ZSTD), and point measurement correlation (Pt.Measure Corr) [2,3,17]. The criteria were as follows: 1) an outfit MNSQ value of 0.5 to 1.5, a ZSTD outfit value of −2 to 2, and 3) a PT. Measure Corr value of 0.4 to 0.85 (Bond & Fox, 2015; Boone et al., 2014; Linacre, 2024). Specifically, in this study, misfit items were categorized based on three main indicators of the Rasch model.

An item is categorized as a good fit if the outfit MNSQ value is in the range of 0.5 to 1.5, the ZSTD value is −2.0 to 2.0, and the Pt. Measure Corr is 0.40 to 0.85 [2,6,17]. An item is categorized as almost misfit if the outfit MNSQ value is 1.3 to 1.5, the ZSTD value is close to ±2.0, and the Pt. Measure Corr is outside the ideal range but not extreme [3,18]. An item is categorized as moderate misfit if the MNSQ value is of 1.5 to 2.0, ZSTD is at ±2.0 to ±3.0, and Pt. Measure Corr is very low (< 0.30) or too high (> 0.90). Items are categorized as severe misfit if the MNSQ value is > 2.0, ZSTD is ≥ ± 3.0, and Pt. Measure Corr is negative [3,17–19]. Furthermore, items are categorized as overfit if the MNSQ value is < 0.5, ZSTD is <−2.0, and Pt. Measure Corr is > 0.90 [2,17,18]. In addition to assessing misfit items, a person fit analysis was also conducted to determine whether participants' response pattern was consistent with the estimated ability according to the Rasch model. This evaluation used the same metrics as the project fit evaluation: MNSQ-Outfit, ZSTD-Outfit, and Pt. Measure Corr. The acceptance criteria were the same as those used in the test project suitability evaluation [2,17,18].

Unidimensionality Analysis

This measuring tool was checked for single-dimension to ensure that it considered only one main hidden idea, matching the key rules of the Rasch modelling [17,20]. Principal component analysis (PCA) was used on the leftovers; change not caused by the main dimension was used to find any possible extra dimensions [2,5,17,20]. There are two key data points here. The first is the variance explained by the Rasch model measure in its original form is expected to be ≥ 50%, with the principal dimension dominating the variance. Second, the residual eigenvalu e of the first contrast was < 3.0, and the secondary dimension was not large enough to distort the measurement [4,21]. This test was interpreted in terms of determining whether the instrument fulfilled the unidimensionality assumption and provided valid and consistent measurement concerning a particular construct.

Differential Item Functioning DIF

According to Hope [10], DIF analysis is used to assess the balance and fairness of a test between respondent groups, identify items that have potential for bias, and increase the construct validity and overall fairness of the measurement instrument. Furthermore, according to Linacre [17] and Sumintono and Widhiarso [22], DIF contrast values can be categorized as follows: 1) no DIF (negligible) if the absolute value of DIF contrast is < 0.43 logits; 2) moderate DIF if it is in the range of 0.43–0.99 logits; and 3) substantial DIF if the absolute value of DIF contrast is ≥ 1.00 logits. The substantial DIF category is further subdivided internally into moderately large DIF if the absolute value of the DIF contrast is in the range of 1.00–1.49 logits and large DIF if the absolute value of the DIF contrast is ≥ 1.50 logits [8]. An item has significant DIF if it meets two main requirements: an absolute value of DIF contrast ≥ 0.64 and p-value < 0.05 [17,22]. If these two criteria are met, then the item is considered to have significant bias and needs to be followed up through further evaluation of the wording and context of the question. This classification ensures a comprehensive interpretation on DIF that takes into account not only effect size but also statistical significance thus making results from DIF analysis informative as well as relevant when used for measurement fairness evaluation between study programs.

All analysis results were then interpreted based on modern measurement theory and Rasch analysis guidelines. The results of statistical fit, unidimensionality, and DIF were synthesized to conclude the overall quality of the instrument, while also providing recommendations for improvements for the development of Applied Mathematics instruments at the Polytechnic.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 01 January 2026
DOI: 10.47191/ijcsrr/V9-i1-16, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

RESULTS AND DISCUSSION

This study did a thorough quality check on the Applied Mathematics FSE test by using the Rasch Model to look at reliability, separation index, item fit, unidimensionality, and DIF. This step-by-step method is meant to make sure that the test is not only consistent and construct valid, but also fair to use in all study programs.

Results

Reliability and Separation Index

Table 1 shows the results of the analysis using the Rasch model for the Applied Mathematics FSE data conducted on 206 students with 40 questions using the Rasch model methodology on Winsteps 5.9 software.

Table 1. Summary of the Statistics of Applied Mathematics FSE Data

Statistics	Person (Non-extreme)	Person (all)	Item
Number of respondents/item	205	206	40
Mean total score	26.0	25.9	133.2
Mean ability/item (logit)	1.32	1.29	0.00
Standard deviation	2.72	2.76	1.60
MNSQ infit (mean)	0.95	-	0.96
MNSQ Outfit (mean)	1.25	-	1.26
ZSTD Infit (mean)	−0.05		−0.34
ZSTD Outfit (mean)	0.19	-	0.20
Logit range	−4.46–5.38	−5.70–5.38	−2.08–6.22
Reliability	0.92	0.92	0.97
Separation index	3.46	3.47	6.18
Score metric correlation	0.98	0.98	1.00
Cronbach's alpha (KR-20)	0.98		
Number of respondents (extreme)	1 (0.5%)	1 (0.5%)	

The Rasch model summary statistics validated the strong psychometric characteristics of the applied mathematics FSE test. Out of the 206 students who participated in this test, 205 were considered non-extreme as only one (0.5%) scored an extreme result. The typical student earned a raw score of 26 out of 40 questions with an average estimated ability of 1.29 logits, standard deviation (SD) = 2.76 ranging from −5.70 to 5.38 logits; hence, there is great variance among students' abilities. In line with Rasch's prescription for item difficulty averaging at zero logit, SD = 1.60, range = −2.08 – 6.22), this shows that there are enough different degrees of difficulty represented by items, including easy and hard ones, on the test. Goodness-of-fit statistics demonstrate that student response patterns and item functioning align well with the hopefulness of the Rasch model. The average respondent MNSQ infit score was 0.95, and the corresponding outfit MNSQ score was 1.25. At the item level, the average MNSQ infit and outfit scores were 0.96 and 1.26, respectively. The mean ZSTD results were approximately zero at the respondent and item levels, indicating no aggregate deviation from the model-implied fit requirements. Person reliability measured 0.92 accompanied by a separation index value of 3.46, indicating that this test can identify not less than four strata in an examinee population. Item reliability was as high as 0.97, showing six classes through a separation index formula based on item difficulties, thereby supporting the hierarchy stability among the tested items. The inside steadiness of the tool, as seen by Cronbach's alpha (KR-20), is also very high, 0.98, which falls in the very high class, showing very good inside steadiness of the tool [3].

Item Suitability

The item suitability analysis with the Rasch model was conducted using the misfit order output from Winsteps 5.9 and is summarized in Table 2.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 01 January 2026
DOI: 10.47191/ijcsrr/V9-i1-16, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

Table 2. Summary of the Misfit Order Analysis

No.	Item Code	Measure (Logit)	MNSQ Infit	Outfit MNSQ	ZSTD	PT. Meas. Corr	Status
1	B31	−2.02	1.47	5.65	3.79	0.48	Severe misfit
2	B14	−2.08	1.09	3.97	2.87	0.58	Severe misfit
3	B34	1.81	1.61	2.74	3.27	0.60	Moderate misfit
4	B36	1.90	1.58	2.07	2.22	0.59	Moderate misfit
5	B6	0.49	0.38	0.19	−3.73	0.89	Moderate misfit
6	B5	1.63	0.96	1.96	2.22	0.74	Almost misfit
7	B32	−0.68	1.54	1.91	1.67	0.63	Needs review
8	B33	−1.95	1.17	1.89	1.30	0.59	Needs review
9	B2	−0.09	1.06	1.82	1.74	0.74	Needs review

Most of the students did well on the applied mathematics FSE. The statistical indicators of fit and completeness of the MNSQ, ZSTD, and point-measure correlations mostly fell within the ranges that would be modeled as tolerable by Rasch. This instrument's items have consistently demonstrated their ability to distinguish test takers across various ability levels. The Rasch model's basic measurement principles, such as unidimensionality, consistency of item logit with ability, and internal validity, have been met. However, some items still showed signs of mild, moderate, or severe misfit. Items classified as highly misfit were predicted not to represent the same construct, or participants' responses deviated from model predictions. Moderately misfit items were most likely because of poorly functioning distractors or unclear question wording. Some items have slight discrepancies even though they are within tolerable limits; however, they need to be fixed to ensure that the quality of the test is optimal. Thus, the items exhibiting this level of discrepancy should be validated more qualitatively expertly assessed or through participant interviews—to determine if improvements or replacements are required. Most items have fulfilled the requirements of Rasch modeling, resulting in high construct validity for this FSE test in applied mathematics. An test with high construct validity can maintain a consistent and accurate measurement of ability in applied mathematics [2,3,6,17,18]wever, some items need to be revised to ensure that the instrument remains valid and reliable whenever it is used for subsequent FSE.

Unidimensionality

The results of the PCA are given in Table 3.

Table 3. Summary of the Unidimensionality Tests

Variance Components	Eigenvalue	Percentage Observed	Percentage Expected
Total raw variance in observations	110.7713	100.0%	100.0%
Variance is explained by measures	70.7713	63.9%	62.5%
• by persons	50.7008	45.8%	44.8%
• by items	20.0705	18.1%	17.7%
Unexplained variance	40.0000	36.1%	37.5%
• Unexplained variance in the 1st contrast	2.5280	2.3%	6.3%
• Unexplained variance in the 2nd contrast	2.0527	1.9%	5.1%
• Unexplained variance in the 3rd contrast	1.7257	1.6%	4.3%
• Unexplained variance in the 4th contrast	1.7090	1.5%	4.3%
• Unexplained variance in the 5th contrast	1.6677	1.5%	4.2%
Essential unidimensionality (Rasch)		88.0%	

The total raw variance of the observations was 110.77 eigenvalue units (Table 3), of which 70.77 units were successfully explained by the Rasch model, with a contribution from respondents at 45.8% and test items at 18.1%. The raw variance explained by the measures is 63.9%, whereas the unexplained variance in the first contrast expresses only 2.3% of the total variance. No

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 01 January 2026
DOI: 10.47191/ijcsrr/V9-i1-16, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

contrast dimension has any unexplained variance exceeding 15%; hence, no dominant secondary dimension exists [9,23]. Essential unidimensionality is calculated to be as high as 88%; therefore, most of the unexplained variance remains on the direction path of the primary dimension, which is consistent with the unidimensionality criteria recommended in Rasch [24]. The residuals revealed several items with substantial positive and negative loadings. The specific values for items with the largest positive loadings were B4 (0.58), B8 (0.51), B9 (0.52), and B6 (0.48). The largest negative loadings were for B3 (−0.36), B34 (−0.34), and B32 (−0.33). The revised clusters' correlations compared with baseline were very high; the attenuated Pearson correlation was close to or equal to 1.00—consistency of measurement across minor dimensions [5,18].

DIF Analysis Between Groups

The DIF analysis showed that a number of items indicated DIF. A summary of the results is presented in Tabel 4.

Table 4. Summary of DIF Analysis Results

No.	Item Code	Study Program Pair	DIF Contrast (logit)	p-value	Klasifikasi DIF	Interpretation (Who has more difficulty
1	B5	(1) vs (2)	-1.17	.0139	moderate large DIF	(2) more difficult than (1)
2	B15	(1) vs (2)	+1.49	.0035	moderate large DIF	(1) more difficult than (2)
3	B16	(1) vs (3)	-2.47	.0077	large DIF	(3) more difficult than (1)
4	B17	(1) vs (2)	-1.40	.0073	moderate large DIF	(2) lebih kesulitan dari (1)
5	B21	(1) vs (2)	-1.32	.0065	moderate large DIF	(2) more difficult than (1)
6	B34	(1) vs (2)	-1.46	.0019	moderate large DIF	(2) more difficult than (1)
7	B36	(1) vs (2)	-1.96	.0001	large DIF	(2) more difficult than 2
8	B39	(1) vs (2)	+1.57	.0009	large DIF	(1) more difficult than (2)

Note: (1) students in the Mechanical Engineering Study Program, (2) students in Refrigeration and Air Conditioning Engineering, and (3) students in Utility Engineering Technology.

The summary of the DIF test in Table 4 shows that there are eight items (B5, B15, B16, B17, B21, B34, B36, and B39) of comparison pairs between PS that show significant and substantial DIF (absolute value of DIF contrast ≥ 0.64 and p-value < 0.05) categorized as substantial DIF. Three items show a large DIF, and five items show a moderate-large DIF. In the Applied Mathematics FSE test reviewed in this content, 80% of the questions are general in nature and do not contain technical contexts that benefit or disadvantage certain groups. The remaining 20% tend to contain technical contexts that potentially benefit certain groups with similar abilities. For example, items B17, B21, and B34 are more difficult for students from the Refrigeration and Air Conditioning Engineering Study Program compared to Mechanical Engineering because they contain machine mechanics contexts that are not commonly studied in their programs [25]. Similarly, items B15 and B39 were more difficult for participants from the Mechanical Engineering Study Program compared to those from the Utility Engineering Technology Study Program or the Refrigeration and Air Conditioning Study Program. This indicates that the test does not purely measure applied mathematics skills but also assesses specific technical knowledge, thus threatening construct validity [2]. This condition is inconsistent with the recommendation by Biggs and Tang [26] that in the framework of assessing learning outcomes, assessments should reflect basic competencies without biasing specific disciplinary backgrounds. Finding and fixing DIF is very important in criterion-based tests [11,27]. This result aligns with research indicating that minor alterations in context or question phrasing can elicit substantial DIF [10,11,27].

Discussion

The study thoroughly evaluated the quality of an applied mathematics FSE utilizing reliability analysis, item discrimination and goodness-of-fit analyses, unidimensionality analysis, and DIF analysis employing the Rasch model. The systematic analyses aimed to evaluate the consistency and validity of the exam items, as well as their equity across various degree programs. Results from reliability analysis together with item discrimination analysis indicated that the exam items possessed high measurement consistency.

The reliability and item discrimination analysis results also support the tool's good internal consistency. High person reliability means that the tool always measures students' skills in applied math, and reasonable person separation means that the tool can put

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 01 January 2026
DOI: 10.47191/ijcsrr/V9-i1-16, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

students into different performance levels. This finding is critical for vocational education because assessment tools need to report not only final grades but also differences in student learning outcomes effectively [2,3].

In terms of test items, high item reliability and item separation index values indicate that the estimated item difficulty level is stable and does not depend excessively on sample characteristics. This indicates that the test items are proportionally distributed over an adequate range of difficulty, thus supporting the instrument's function as a reliable measuring tool [19,22,23].

Additionally, the item fit examination showed that most items was Infit and Outfit Mean Square values which were within the acceptable range of values. This finding shows that the students' answers to the questions were what the Rasch Model expected. The questions did a good job of measuring the skills that were meant to be measured. Although there were several items with a tendency toward mild misfitness, this condition is still tolerable and is more appropriately addressed through editorial revision or adjustment of the question context rather than outright deletion. In the context of vocational education, mild misfit is often related to the use of certain technical contexts that are not completely uniform across study programs [2].

The adequate reliability and item fit results provide a strong basis for testing the instrument's unidimensionality, a fundamental assumption in the Rasch Model. Table 23 principal component analysis of Rasch residuals shows that the model explained 63.9% of the variance. This is more than the 40% minimum limit that Rasch modeling suggests.

Furthermore, the unexplained variance values in the first comparison indicate that there are no other major latent dimensions. The essential unidimensionality value of 88% further confirms that most items contribute consistently in measuring one main construct, namely students' Applied Mathematics ability. This finding indicates that even though the test was used on students from three different study programs, its measurement structure remains stable and is not fragmented by differences in academic background [17,28].

Meeting the unidimensionality assumption allows for more valid interpretations of further analyses, particularly the DIF analysis. The DIF analysis results in Table 4 indicate that the test is generally fair, as indicated by the majority of items not showing significant differences in item function across study programs. This finding reinforces the unidimensionality results, as construct stability is also reflected in the equivalence of item functions across groups.

However, several items were identified as showing significant DIF within the Substantial DIF category. Items B5, B17, B21, and B34 exhibited significant DIF within the moderate-large DIF range, signifying disparities in success rates among study program groups. Meanwhile, items B16, B36, and B39 showed substantial DIF within the large DIF range. It is important to recognize that there is substantial amount of DIF across these eight items. Even when that aren't many of those items with strong biases may influence how fair the measurement and interpretation of test results are actually [11,29].

As a whole, the Application of Mathematics FSE test has great psychometric qualities and is a good tool for evaluating learning outcomes. That depends on the combination of reliability, item suitability, unidimensionality, and DIF results. But the DIF results show that just meeting the standards for unidimensionality and reliability for internal consistency is not enough to say that decisions are fair. Consequently, DIF analysis is an essential component of instrument assessment, specifically for vocational education according to the framework of outcome-based education, where assessments across study programs need instruments that are not only valid and reliable but also fair and equitable.

These findings provide a basis for lecturers and instrument developers to refine potentially biased items to make cross-program assessments fair and accountable. Theoretically, these results reinforce the urgency of simultaneously testing unidimensionality and DIF in Rasch Model-based instruments. Therefore, this article has contributed to the development of valid, reliable, and fair assessment practices in polytechnics.

The study implies several important aspects for developing assessment and learning practices within vocational education environments: 1) asking developers to compile test items by emphasizing the principle of construct equivalence, 2) modifying or harmonizing course LOs so that they become representative toward the needs of all SPs under the same department, 3) Academic quality assurance parties truly need to periodically audit assessment instruments using a quantitative approach based on IRT to ensure that the exam measures competency fairly and accurately. 4) Assessments require further refinement through multi-party involvement in item construction, context validity testing, and systematic reviews based on Rasch model data. Finally, regular training is needed for lecturers on question construction techniques, item fit and DIF analysis, and usage of assessment data for continuous improvement.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 01 January 2026
DOI: 10.47191/ijcsrr/V9-i1-16, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

CONCLUSION

Overall, this analysis concludes that the FSE applied mathematics instrument has broadly met the basic requirements of measurement. More specifically, for the instrument: (1) Evidence supports unidimensionality. The variance explained by the model was 63.9%, with a low eigenvalue on residual contrast (2.528), therefore indicating that this instrument measures one main construct and can be used to report scores based on a single scale; (2) There was no DIF found in most items; hence, it is relatively fair across study programs in measuring student ability. However, items B5, B15, B16, B17, B21, B34, B36, and B39 exhibited significant DIF, potentially introducing bias between study program groups. These items need to be revised or replaced so that no group is disadvantaged; (3) Based on the results of unidimensionality and DIF analysis, it can be seen that Applied Mathematics FSE has good psychometric characteristics but still needs refinement on the items that contain bias. Therefore, this research becomes an empirical consideration for stating that although the test is construct valid and relatively fair, Eight items still urgently need revisions to make the assessment more objective and apply equity in line with vocational education standards.

RECOMMENDATIONS

These results helped to beef up the theoretical basis of measurement in mathematics education and framed practically how valid and reliable assessment tools can be developed. It is very handy for teachers, lecturers, and researchers at any level who want to assess applied mathematical competence. Instrument development is a continuous process that requires further refinement and validation [30]

ACKNOWLEDGEMENTS

Thanks are due to the Mechanical Engineering Study Program, PNB, and all teaching staff who provided kind support, facilities, and guidance in carrying out this research. We extend our gratitude to the study participants and other individuals who contributed to the completion of this work. The authors wish to thank the editors and anonymous reviewers of IJCSRR for their valuable comments and suggestions, which led to improvements in the quality of this manuscript. This study belongs here under the Knowledge Development umbrella, with a specific psychometric evaluation. Public, commercial, or not-for-profit funding agencies have not specifically funded this study.

REFERENCES

1. Hake, R. R. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys.* 66, 64–74 (1998).
2. Bond, T. G. & Fox, C. M. *Applying the Rasch model: Fundamental measurement in the human sciences,*. (Routledge, 2015). doi:https://doi.org/10.1111/j.1745-3984.2003.tb01103.x.
3. Boone, W. J., Staver, J. R. & Yale, M. S. *Rasch Analysis in the human sciences*. (Springer., 2014). doi:10.1007/978-94-007-6857-4.
4. Tennant, A., McKenna, S. P. & Hagell, P. Application of Rasch analysis in the development and application of quality of life instruments. *Value Heal.* 7, 22–26 (2004).
5. Wright, B. D. & Stone, M. H. *Measurement essentials*. *Measurement* (WIDE RANGE, INC, 1999).
6. Linacre, J. M. What do infit and outfit, mean-square and standardized mean? *Rasch Meas. Trans.* 6, 878 (2012).
7. Tennant, A. & Küçükdeveci, A. A. Application of the Rasch measurement model in rehabilitation research and practice: Early developments, current practice, and future challenges. *Front. Rehabil. Sci.* 4, 01–17 (2023).
8. Dorans, N. J. & Holland, P. W. *DIF Detection and Description: Mantel-Haenszel and Standardization*. *Differential Item Functioning* (N.J: Lawrence Erlbaum Associates., 1993).
9. Linacre, J. M., Boone, W., Green, R., Trevor, G. & Christine, M. A User's Guide to WINSTEPS® / MINISTEP Rasch-model computer programs. 2–4 (2023).
10. Hope, D., Adamson, K., McManus, I. C., Chis, L. & Elder, A. Using differential item functioning to evaluate potential bias in a high stakes postgraduate knowledge based assessment. *BMC Med. Educ.* 18, 1–7 (2018).
11. Zumbo, B. D. Psychometric methods for investigating DIF and test bias during test adaptation across languages and cultures. 1–85 (2006).

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 01 January 2026
DOI: 10.47191/ijcsrr/V9-i1-16, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

12. Holland, P. W. & Wainer, H. *Differential item functioning*. (Routledge., 2012).
13. Creswell, J. W. & Creswell, J. D. *Research design qualitative, quantitative, and mixed methods approaches*. (SAGE Publications, Inc., 2017).
14. Kesmodel, U. S. Cross-sectional studies – what are they good for? *Acta Obstet. Gynecol. Scand.* 97, 388–393 (2018).
15. Aiken, L. R. Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educ. Psychol. Meas.* 45, 131–142 (1985).
16. Wright, B. D. & Masters, G. N. *Rating scale analysis item response modeling approach*. (Mesa Press, 1982).
17. Linacre, J. M. Reasonable mean-square fit values revisited. *Rasch Meas. Trans.* 36, 1–6 (2024).
18. Wright, B. D. & Linacre, J. M. Reasonable mean-square fit values. *Rasch Meas. Trans.* 8, 370 (1994).
19. Linacre, J. M. Optimizing rating scale category effectiveness optimizing rating scale category effectiveness university of Chicago. *J. Appl. Meas.* 3, 85–106 (2002).
20. Linacre, J. M. Data Variance Explained by Rasch Measures. *Rasch Meas. Trans.* 20, 1045 (2006).
21. Supriatna, M., Suryana, D., Anzhali, M. N. & Noor, A. M. Evaluating career self-awareness instruments for victims of violence using the Rasch model. *BMC Psychol.* 13, 1–13 (2025).
22. Sumintono, B. & Widhiarso, W. *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. (Trim Komunikata, 2015).
23. Linacre, J. M. Predicting responses from rasch measures. *J. Appl. Meas.* 11, 1–10 (2010).
24. Bond, T. G., Yan, Z. & Heene, M. *Applying the Rasch model fundamental measurement in the human sciences*. (Routledge, 2020). doi:https://doi.org/10.4324/9780429030499.
25. Hambleton, R. K., Swaminathan, H. & Rogers, H. J. Fundamentals of item response theory. *Am. Sosiological Assoc.* 21, 289–290 (1992).
26. Biggs, J. & Tang, C. *Teaching for quality learning at university. What the student does (4th Edn.)*. *Innovations in Education and Teaching International* vol. 50 (SRHE and Open University Press, 2011).
27. Zumbo, B. D. *A handbook on the theory and methods of differential item functioning (DIF)*. *Ottawa: National Defense Headquarters* (Directorate of Human Resources Research and Evaluation Department of National Defen, 1999).
28. Linacre, J. M. Rasch measurement transactions: Teaching Rasch measurement. *Autumn.* 31, 1630–1642 (2017).
29. Zumbo, B. D. & Gelin, M. N. A matter of test bBias in educational policy research: Bringing the context into picture by iInvestigating sociological / community moderated (or mediated) test and item bias. *J. Educ. Res. Policy Stud.* 5, 1–23 (2005).
30. Engelhard, G. & Wang, J. *Rasch Models for Solving Measurement Problems: Invariant Measurement in the Social Sciences*. (SAGE Publications, Inc., 2022). doi:10.4135/9781071878675.

Cite this Article: Darma, I.K., Herdiansyah, A., Rofi'I, A. (2026). Unidimensionality Analysis and Differential Item Functioning in the Applied Mathematics Final Examination Test of Politeknik Negeri Bali Students Using the Rasch Model. International Journal of Current Science Research and Review, 9(1), pp. 130-138. DOI: https://doi.org/10.47191/ijcsrr/V9-i1-16