INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

OPEN ACCESS JOURNAL

Vol 14, Issue 3, (2025) E-ISSN: 2226-6348

A Rasch Model Analysis of the Suitability of Exam Items in DBS10042 Engineering Science

Geetha Subramaniam, Muhammad Syafiq Bin Abdul Ghafar, Intan Noor Dzalika Binti Azis

Mathematics, Science and Computer Department, Polytechnic Sultan Idris Shah, Sungai Lang, 45100 Sungai Ayer Tawar, Selangor, Malaysia

Email: geethasubramaniam@psis.edu.my

DOI Link: http://dx.doi.org/10.6007/IJARPED/v14-i3/25932

Published Online: 25 August 2025

Abstract

In educational settings, the Rasch Measurement Model is widely used to ensure that assessment instruments produce trustworthy and consistent results. This study applied the Rasch model to analyse the final examination items of the DBS10042 Engineering Science course offered at a Malaysian polytechnic. A total of 240 first-semester engineering students participated in the study. The analysis focused on item fit statistics, item difficulty, and point-measure correlations to determine how well each question aligned with students' abilities and the intended learning outcomes. Results indicated that most items fell within acceptable fit ranges, confirming their appropriateness for assessing student performance. However, a few items exhibited misfit, suggesting issues such as ambiguity, misalignment with cognitive levels, or excessive difficulty. The study also revealed strong person-item correlations for most items, supporting their diagnostic value in differentiating student proficiency. Overall, the findings highlight the effectiveness of using the Rasch Measurement Model to ensure high quality, fair, and reliable assessments in technical and vocational education contexts.

Keywords: Validity, Reliability, Questions, Examination, Rasch

Introduction

Assessing learning outcomes is crucial for monitoring students' learning processes, progress, achievements, and ongoing improvement. It provides valuable insights into the effectiveness of teaching strategies, the depth of student understanding, and the overall quality of the educational experience(Abdullah et al., 2012; Hope et al., 2024; Osmin & Zainuddin, 2021). Effective evaluation requires accurate evidence that reflects the extent of students' mastery of learning outcomes, which can enhance both motivation and achievement. In Malaysia's educational framework, science is a core subject taught at all levels, from primary schools to universities (Saleh, 2021; Yusop et al., 2022). Within the Technical and Vocational Education and Training (TVET) system, particularly in polytechnics, Engineering Science plays a vital role in equipping students with foundational knowledge and essential technical skills (Amin et al., 2023; Mamat, 2023).This subject is critical for preparing students to address real-world engineering challenges. At the polytechnic level, Engineering Science courses are mandatory for first-semester engineering students and cover topics such as physical quantities and measurement, linear motion, forces, fluids, and thermodynamics.

Given its significance, it is essential to thoroughly evaluate learning outcomes in Engineering Science to assess students' understanding and practical application of key concepts(Hernánez-Suarez et al., 2021; Nazry Ali et al., 2022; Ozkan & Topsakal, 2020). Final semester exams serve as the primary tool for gauging students' comprehension and proficiency. These exams provide valuable insights into learners' strengths and weaknesses, guiding both academic improvement and curriculum refinement (Azizah et al., 2022; Khalid et al., 2024; Lohgheswary et al., 2019; Lohgheswary, Salmaliza, et al., 2022).

However, the effectiveness of such evaluations depends on the quality of the exam items. Well-constructed questions can accurately measure student performance, while poorly designed items may obscure true proficiency levels (Lohgheswary, Diana, et al., 2022; Mamat, 2023; Octaviana et al., 2022). Despite the importance of item analysis, a method used to evaluate the performance of individual test items, it is often underutilized. This underuse may stem from time constraints, a lack of awareness, or insufficient training among educators (Herrmann-Abell et al., 2018; Istiyono et al., 2020; Lohgheswary, Lun, et al., 2022). To address this gap, it is important to employ more rigorous analytical techniques to enhance the validity and reliability of assessments. One such method is the Rasch Measurement Model. Renowned for its ability to analyze test data at a granular level, the Rasch Model evaluates item difficulty, discrimination, and fit, offering detailed diagnostic information (Lohgheswary et al., 2019; Lohgheswary, Lun, et al., 2022; Neumann et al., 2011). Unlike traditional item analysis, the Rasch Model provides insights into both individual item performance and overall test scale reliability, thus yielding a more accurate representation of students' abilities and learning outcomes. This study applies the Rasch Measurement Model to examine the suitability of DBS10042 Engineering Science exam items. Its primary objective is to determine whether these test items meet established standards for validity and reliability, ensuring they effectively measure students' mastery of course content.

Literature Review

The quality of examination items plays a critical role in ensuring that assessments accurately measure student competencies and support effective instructional decision-making. Prior findings showed that well-designed test items not only assess comprehension but also offer diagnostic information for improving instruction (Azizah et al., 2022; Octaviana et al., 2022). One widely accepted method for evaluating test item quality is item analysis, which assesses reliability, validity, and the overall performance of individual questions. Traditional approaches to item analysis, such as those based on Classical Test Theory (CTT), are widely used to evaluate test reliability and validity (Neumann et al., 2011; Stemler & Naples, 2021; Yulisharyasti et al., 2023). However, CTT is limited by its dependence on sample-specific data and its inability to distinguish between item characteristics and student ability. To overcome these limitations, researchers have increasingly adopted the Rasch Measurement Model (RMM), a probabilistic model under the framework of Item Response Theory (IRT). Rasch analysis enables invariant measurement by independently estimating item difficulty and person ability, thereby offering more precise and generalizable insights into assessment quality (Stemler & Naples, 2021; Wicaksono & Korom, 2023; Yulisharyasti et al., 2023). Throughout Technical and Vocational Education and Training (TVET), the Rasch model has demonstrated its effectiveness in identifying misfitting items, refining assessments, and aligning test content with learning objectives. Despite its advantages, the application of Rasch analysis at the polytechnic level remains limited, particularly for high-stakes assessments such as final examinations. Foundational engineering courses like DBS10042, offered in Malaysian polytechnics, have not been extensively studied using this method. This gap presents an opportunity to strengthen assessment practices in TVET institutions. Therefore, this study employs the Rasch Measurement Model to analyse the final examination items of the DBS10042 Engineering Science course, with the aim of evaluating their validity, reliability, and alignment with student learning outcomes.

Methodology

A quantitative research method was employed in this study to analyse student performance through final examination outcomes in the DBS10042 Engineering Science course. The sampling method used was census sampling, whereby the entire population of students enrolled in the course during the 2024/2025 academic semester was included. A total of 240 first-semester engineering students participated in the study, comprising 10 students from Chemical Engineering, 3 from Electronic and Electrical Engineering, 15 from Mechanical Engineering, and 6 from Civil Engineering. Throughout the semester, students received 14 weeks of instruction and completed multiple course components contributing to their final grade. These components included a test (15%), practical work (20%), a mini project (25%), and a final examination (40%). The final course grade was derived from the weighted aggregate of these components. The research instrument used to assess student performance was the final examination, structured according to Bloom's Taxonomy, specifically targeting the cognitive domains of Remember, Understand, and Apply (Heryani et al., 2021; Lohgheswary, Diana, et al., 2022). The examination comprised four structured subjective questions, each consisting of multiple parts, resulting in a total of 24 items. The exam carried a total of 100 marks and had a duration of two hours. For data organization and analysis, student examination scores were entered into Microsoft Excel and subsequently analysed using the Winstep software, which supports the application of the Rasch Measurement Model. This model was utilized to evaluate the quality of individual test items in terms of difficulty, fit statistics, and alignment with expected cognitive levels.

Findings and Discussion

In the current study, the fit statistics (Infit MNSQ and Outfit MNSQ) and Person-Measure Correlation (PT-MEASURE CORR.) for each item were analysed to assess the validity and reliability of the data obtained from 240 participants. Table 1 presents a summary of these metrics for each item.

Table 1
Summary of Metrics

ITEM	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	OBS%	EXP%
3aii	2.26	8.5	1.98	5.4	.47	47.1	48.5
3bii	2.07	9.9	1.91	8.5	.67	3.8	21.2
4c	2.06	9.7	1.90	8.1	.64	9.2	20.4
3cii	1.81	6.2	1.32	2.0	.44	37.1	40.7
4aii	1.66	5.7	1.48	3.6	.54	37.1	35.7
1cii	1.53	5.4	1.53	5.3	.66	16.7	29.5
2aii	1.36	3.6	1.36	3.3	.54	20.0	31.0
2bii	1.33	3.3	1.16	1.5	.69	25.4	29.4
3ci	1.18	1.9	1.30	2.9	.59	21.7	30.5
1aii	1.16	1.7	1.29	2.6	.46	21.3	30.5
2cii	1.26	2.7	1.21	2.1	.62	24.6	31.0
2bi	.95	-0.5	1.02	0.3	.51	34.2	30.5
1biii	.92	-1.0	.95	-0.6	.47	33.3	30.6
4ai	.54	-6.0	.92	-0.6	.39	48.3	33.7
1bii	.63	-4.7	.74	-2.9	.50	39.2	30.7
4bii	.71	-3.6	.74	-2.8	.57	39.2	30.5
4biii	.71	-3.6	.74	-2.8	.57	39.2	30.5
2ai	.46	-7.6	.72	-3.0	.44	42.1	30.6
1bi	.66	-4.3	.71	-3.2	.52	37.5	30.6
1ai	.36	-9.7	.68	-3.3	.50	52.1	31.2
3ai	.43	-7.8	.66	-3.1	.57	57.1	34.1
4bi	.66	-4.5	.65	-4.3	.68	41.3	31.1
3bi	.48	-7.2	.63	-4.3	.64	50.4	30.7
4bii	.60	-5.5	.62	-5.1	.66	43.8	29.3
1ci	.49	-7.0	.57	-5.2	.58	49.6	30.9
2ci	.45	-7.9	.53	-6.0	.62	47.9	30.4

The fit statistics for each item, specifically the Infit and Outfit Mean Square (MNSQ) values, were analysed to assess how well each item conformed to the expectations of the Rasch measurement model. Infit MNSQ values in this study ranged from 0.36 to 2.26, while Outfit MNSQ values ranged from 0.53 to 1.98. According to established Rasch guidelines, MNSQ values between 0.5 and 1.5 are considered acceptable for productive measurement(Azizah et al., 2022; Rahim & Haryanto, 2021). Items falling outside this range may indicate misfit either due to unpredictability (values above 1.5) or redundancy (values below 0.5). In this dataset, several items exceeded the upper threshold, suggesting underfit. Notably, Item 3aii (Infit MNSQ = 2.26; Outfit MNSQ = 1.98) and Item 3bii (Infit MNSQ = 2.07) displayed the most significant misfit, implying that student responses to these items were more erratic than the model predicted. These items may have been poorly worded, conceptually complex, or misaligned with students' skill levels. Conversely, several items demonstrated strong fit. Item 2bi (Infit MNSQ = 0.95; Outfit MNSQ = 1.02) and Item 1biii (Infit MNSQ = 0.92; Outfit MNSQ = 0.95) were well within the acceptable range, suggesting these items performed predictably and appropriately targeted student ability levels. Items with low MNSQ values, such as Item 1ai (Infit MNSQ = 0.36; Outfit MNSQ = 0.68), may be too easy or redundant, contributing less new information to the measurement of ability. Overall, the distribution of MNSQ values indicates that while many items functioned well, a few require revision to improve fit and diagnostic usefulness.

The Point-Measure Correlation (PT-MEASURE CORR.) values ranged from 0.39 to 0.69, reflecting moderate to strong positive correlations between individual item responses and the overall student ability estimate. This indicates that most items positively contributed to differentiating students across ability levels. The highest correlations were observed for Item 2bii (0.69), Item 4bi (0.68), and Item 4bii (0.66), suggesting these items closely aligned with the latent trait being measured and effectively distinguished between more and less proficient students. On the lower end, Item 4ai (0.39) and Item 3cii (0.44) had weaker correlations, which may indicate lower item quality or reduced alignment with the intended construct. These items should be further reviewed for clarity, content relevance, or alignment with learning outcomes. The Observed (OBS%) and Expected (EXP%) match percentages represent how closely actual student responses matched model predictions. Ideally, observed and expected values should be relatively close, indicating that the item functioned consistently with Rasch model expectations. For most items, the observed match percentages were generally consistent with expected values. For example, Item 3aii had an OBS% of 47.1% and an EXP% of 48.5%, suggesting good alignment. However, some items demonstrated large deviations. Item 3ai had an observed match of 57.1% versus an expected 34.1%, and Item 1ai showed 52.1% observed versus 31.2% expected.

Such large gaps may suggest that these items were unusually easy, overly familiar, or possibly misinterpreted in ways that led to more uniform correct responses than predicted. These discrepancies do not automatically indicate poor item quality but highlight the need for careful examination of how item difficulty and clarity may affect response patterns. Items with high observed percentages but lower expected values might benefit from rewording or cognitive level recalibration to better reflect a balanced assessment profile (Herrmann-Abell et al., 2018; Octaviana et al., 2022). Based on these findings, the application of the Rasch Measurement Model in this study provided important insights into the quality and functionality of the DBS10042 Engineering Science examination items. The findings indicated that most items fell within the acceptable Infit and Outfit Mean Square (MNSQ) range of 0.5 to 1.5, which is widely regarded as optimal for ensuring productive measurement.

This suggests that most of the items were well-targeted and functioned as intended in assessing students' understanding of course content. However, several items such as 3aii, 3bii, and 4c displayed significant misfit, with Infit MNSQ values exceeding 2.0. These results align with prior findings that reported misfitting items could compromise the measurement accuracy by introducing noise or failing to align with students' ability levels. These misfits may be attributable to ambiguous wording, misalignment with Bloom's cognitive domains, or content that was either too complex or not covered adequately in instruction. In contrast, well-performing items like 2bi, 1biii, and 2bii demonstrated MNSQ values within the ideal range and high Point-Measure Correlations (up to 0.69), indicating strong alignment with students' latent traits. These findings are consistent with prior findings that emphasized the importance of well-calibrated items in improving test reliability and providing accurate diagnostic information for both educators and learners (Azizah et al., 2022; Nazry Ali et al., 2022). The analysis of Point-Measure Correlations (PT-MEASURE CORR.) also supports the overall quality of the assessment. The majority of items had correlations above 0.50, indicating moderate to strong relationships between item performance and student ability. This mirrors previous studies that demonstrated that items with strong person-measure correlations contribute more meaningfully to evaluating learning outcomes and guiding instructional refinement (Boone, 2016; Hope et al., 2024; Rustam et al., 2017).

Furthermore, discrepancies between observed and expected match percentages were notable in a few items, such as 1ai and 3ai, where observed matches were significantly higher than expected. These results may suggest that some items were too easy or led to patterned responses, potentially reducing their discriminative power. Previous studies also highlighted similar issues, noting that substantial gaps between observed and expected scores can point to content familiarity or surface-level memorization rather than deep understanding (Azizah et al., 2022; Hadi & Lestari, 2024; Yulisharyasti et al., 2023). Moreover, this study affirms the pedagogical value of incorporating Bloom's Taxonomy into item design. As noted by previous findings, aligning questions to different cognitive levels helps ensure a comprehensive evaluation of students' abilities from recall and understanding to application. However, the misfit of some items also implies a potential mismatch between the intended and actual cognitive level, suggesting the need for more rigorous item validation during test construction. The Rasch Model, as supported by previous educational research, proves to be an effective tool for enhancing the validity, reliability, and instructional alignment of high-stakes assessments in technical and vocational education settings.

Conclusion

This study utilized Rasch analysis to assess the effectiveness and appropriateness of the final examination items in the DBS10042 Engineering Science course at the polytechnic level. The analysis revealed that while most of the exam items demonstrated acceptable fit within Rasch parameters, a number of items such as 3aii, 3bii, and 4cexhibited significant misfit. These results suggest that some items may have been overly difficult, ambiguous, or misaligned with the measured construct, thereby compromising their validity. The person-measure correlations for most items were moderate to strong, indicating that the items generally contributed meaningfully to differentiating between students of varying ability levels. Additionally, the comparison of observed and expected match percentages identified several items with large discrepancies, suggesting a need for further review and potential revision of those questions. Overall, the findings underscore the value of using Rasch analysis in educational assessment to enhance test validity and reliability. This model provides detailed diagnostic information that can inform item revision and support the development of fairer, more effective assessments.

Contribution of Knowledge

This study significantly contributes to the expanding body of literature on psychometric evaluation and quality assurance in educational assessment, particularly in the realm of Technical and Vocational Education and Training (TVET). Theoretically, it extends the application of the Rasch Measurement Model by illustrating its effectiveness in evaluating the construct validity, reliability, and diagnostic strength of high-stakes examination items in foundational engineering science courses. The analysis provides empirical evidence supporting the model's capacity to detect item misfit, assess alignment with cognitive domains, and enhance the interpretative power of test scores (Lohgheswary, Lun, et al., 2022; Zafrullah et al., 2023). Contextually, this research addresses a notable gap within the

Malaysian polytechnic environment, where formal item-level validation of assessment instruments tends to be underutilized. By concentrating on the DBS10042 Engineering Science course, the study offers a practical framework for refining the design and evaluation of assessments in STEM-related TVET programs (Amin et al., 2023; Azmi & Salleh, 2021). The findings are particularly relevant for educators, assessment designers, and policymakers aiming to improve the fairness, transparency, and instructional alignment of learning evaluations. Moreover, this research aligns with Malaysia's broader educational transformation agenda by advocating for evidence-based assessment practices that can be implemented across similar courses and institutions within the polytechnic system.

References

Abdullah, H., Arsad, N., Hashim, F. H., Aziz, N. A., Amin, N., & Ali, S. H. (2012). Evaluation of Students' Achievement in the Final Exam Questions for Microelectronic (KKKL3054) using the Rasch Model. *Procedia - Social and Behavioral Sciences*, *60*(October), 119–123. https://doi.org/10.1016/j.sbspro.2012.09.356

Amin, S. M., Ahmad Suhaimi, S. S., & Nazuri, N. S. (2023). The Present and Future of Malaysian Technical and Vocational Education and Training (TVET). *International Journal of Academic Research in Business and Social Sciences*, *13*(18), 107–117. https://doi.org/10.6007/ijarbss/v13-i18/19952

Azizah, N., Suseno, M., & Hayat, B. (2022). Item analysis of the rasch model items in the final semester exam indonesian language lesson. *World Journal of English Language*, *12*(1), 15–26. https://doi.org/10.5430/wjel.v12n1p15

Azmi, T., & Salleh, D. (2021). a Review on Tvet Curriculum Practices in Malaysia. *International Journal of Education, Psychology and Counseling*, *6*(40), 35–48. https://doi.org/10.35631/ijepc.640003

Boone, W. J. (2016). Rasch analysis for instrument development: Why,when,and how? *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-04-0148

Hadi, T. S., & Lestari, I. (2024). *Application of The Rasch Model in Research Publications : A Bibliometric Analysis*. *4*(May), 229–240.

Hernánez-Suarez, C. A., Gamboa-Suárez, A. A., & Suarez, O. J. (2021). Attitudes towards physics. A study with high school students from the Colombian context. *Journal of Physics: Conference Series*, *2118*(1). https://doi.org/10.1088/1742-6596/2118/1/012019

Herrmann-Abell, C. F., Hardcastle, J., & DeBoer, G. E. (2018). Using Rasch to develop and validate an assessment of students' progress on the energy concept. *Grantee Submission*.

Heryani, T. P., Sinaga, P., & Chandra, D. T. (2021). Analysis mastery of concepts physics on the topics of energy for high school students in distance learning during Covid–19. *Journal of Physics: Conference Series*, *2098*(1). https://doi.org/10.1088/1742-6596/2098/1/012003

Hope, D., Kluth, D., Homer, M., Dewar, A., Jaap, A., & Cameron, H. (2024). *items for multi-site and multi-year assessment*.

Istiyono, E., Dwandaru, W. S. B., Setiawan, R., & Megawati, I. (2020). Developing of computerized adaptive testing to measure physics higher order thinking skills of senior high school students and its feasibility of use. *European Journal of Educational Research*, *9*(1), 91–101. https://doi.org/10.12973/eu-jer.9.1.91

Khalid, I. L., Abdullah, M. N. S., & Mohd Fadzil, H. (2024). A Systematic Review: Digital Learning in STEM Education. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *51*(1), 98–115. https://doi.org/10.37934/araset.51.1.98115

Lohgheswary, N., Diana, A. S. F. N., Preethy, A., & Lun, A. W. (2022). Analysis on numerical analysis final exam questions. *International Journal of Health Sciences*, *April*, 1165–1174. https://doi.org/10.53730/ijhs.v6ns2.5076

Lohgheswary, N., Lun, A. W., & Jedi, A. (2022). Evaluating performance of students in engineering statistics final exam questions. *International Journal of Health Sciences*, *April*, 1002–1011. https://doi.org/10.53730/ijhs.v6ns2.5073

Lohgheswary, N., Preethy, A., & Gobithaasan, R. U. (2019). Analysis on Engineering Mathematics II Final Exam Questions. *International Journal of Advanced Science and Technology*, *28*(8s), 123–129.

Lohgheswary, N., Salmaliza, S., Lun, A. W., Jedi, A., & Nopiah, Z. M. (2022). Evaluating reliability of final exam questions via rasch model. *International Journal of Health Sciences*, *April*, 1065–1074. https://doi.org/10.53730/ijhs.v6ns2.5064

Mamat, F. A. (2023). Fokus Penilaian Alternatif (PALT) Sains Kejuruteraan. *Jurnal Dunia Pendidikan*, *5*(1), 36–40. https://doi.org/10.55057/jdpd.2023.5.1.4

Nazry Ali, M., Ghani, W., Haiman Abdul Rahman, A., & Masrom, M. (2022). The ICT implementation in the TVET teaching and learning environment during the COVID-19 pandemic. *International Journal of Advanced Research in Future Ready Learning and Education*, *27*(1), 43–49.

Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, *33*(10), 1373–1405. https://doi.org/10.1080/09500693.2010.511297

Octaviana, R. I., Anggara, M. B., Jamilah, R., Darmana, A., & Suyanti, R. D. (2022). Analisis Item Soal Kimia SMA Menggunakan Rasch Model. *Orbital: Jurnal Pendidikan Kimia*, *6*(1), 26–37. https://doi.org/10.19109/ojpk.v6i1.12248

Osmin, S. S., & Zainuddin, S. A. (2021). *PROGRAM EDUCATIONAL OBJECTIVES ASSESSMENT FOR DIPLOMA IN INFORMATION TECHNOLOGY (DIGITAL TECHNOLOGY) PROGRAMME : EMPLOYERS SURVEY*. *1*(1), 1–13.

Ozkan, G., & Topsakal, U. U. (2020). Determining Students' Conceptual Understandings of Physics Concepts. *Shanlax International Journal of Education*, *8*(3), 1–5. https://doi.org/10.34293/education.v8i3.2908

Rahim, A., & Haryanto, H. (2021). Journal of Educational Research and Evaluation Implementation of Item Response Theory (IRT) Rasch Model in Quality Analysis of Final Exam Tests in Mathematics Article Info. *Journal of Educational Research and Evaluation*, *10*(2), 57–65. http://journal.unnes.ac.id/sju/index.php/jere

Rustam, M., Rameli, M., Mohd, A., & My, K. P.-A. (2017). *Malaysian School Students' Math Anxiety: Application of Rasch Measurement Model*. *January 2016*.

Saleh, S. (2021). Malaysian students' motivation towards Physics learning. *European Journal of Science and Mathematics Education*, *2*(4), 223–232. https://doi.org/10.30935/scimath/9414

Stemler, S. E., & Naples, A. (2021). Rasch Measurement v. Item Response Theory: Knowing When to Cross the Line. *Practical Assessment, Research and Evaluation*, *26*, 1–16. https://doi.org/10.7275/v2gd-4441

Wicaksono, A. G. C., & Korom, E. (2023). Attitudes towards science in higher education: Validation of questionnaire among science teacher candidates and engineering students in Indonesia. *Heliyon*, *9*(9). https://doi.org/10.1016/j.heliyon.2023.e20023

Yulisharyasti, L., Nurdin, A., Aulia, N., Arfa, F. A. H., & Fadjryani. (2023). Analyzing the Quality of Measurement Instruments of Multiple Choice Questions on Class Xi Economics Material in Public High School 3 Gorontalo Through Classical Test Theory and Rasch Models. *Parameter: Journal of Statistics*, *3*(1), 28–38. https://doi.org/10.22487/27765660.2023.v3.i1.16417

Yusop, S. R. M., Rasul, M. S., Yasin, R. M., Hashim, H. U., & Jalaludin, N. A. (2022). An Assessment Approaches and Learning Outcomes in Technical and Vocational Education: A Systematic Review Using PRISMA. *Sustainability (Switzerland)*, *14*(9). https://doi.org/10.3390/su14095225

Zafrullah, Sa'adatul Ulwiyah, & Nofriyandi. (2023). Rasch Model Analysis on Mathematics Test Instruments: Biblioshiny (1983-2023). *Mathematics Research and Education Journal*, *7*(2), 1–13. https://doi.org/10.25299/mrej.2023.vol7(2).14550