ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Research Article; 2025; 16(2); 59-70

Assessment of Student Performance for Laboratory Application Course Examination Using Rasch Measurement Model

Ahmet Volkan Yüzüak[*] Zekiye Merve Öcal[**]

Abstract

Rasch Measurement Model creation in social science education has quickly spread to other educational domains, such as technical and engineering sciences. This research attempts to move testing and assessment in Laboratory Application in Science Education away from the traditional paradigm and toward the bio-based Rasch Model. Based on Kuhn's description of the purpose of measurement, it is thought that the Learning Performance Measurement System (LPMS), in particular, is compatible with the basic measurement now in use. These cannot be obtained from engineering science or statistics textbooks. Following a paradigm change, the Faculty of Education at one of the higher education institutions in Turkiye has started using the Rasch Model to assess the success of its Course Learning Outcomes (CLO): Decision Support System. The Rasch Model provides an accurate summary of the students' success on a linear scale for measuring by tabulating the students, i.e. Person and Items on a Distribution Map (PIDM). A comparison between the Rasch measurement and the conventional histogram tabulation and simple means reveals that the former provides a more thorough exploratory depth for comprehending difficulties in science education courses. Even with the limited sample size, the students were well-categorized based on their individual cognitive talents; as a result, CLOs were created using Bloom's Taxonomy as the structural basis. It is therefore feasible to apply this methodology to the evaluation of general abilities in professionals or students. As a result, the Rasch model of evaluating an individual's ability adopts a new paradigm.

Keywords: *Rasch Measurement Model, Course Learning Outcomes, Performance Assessment, Laboratory Applications, Science Education*

Introduction

In educational area, it has been implementing Outcome-Based Education (OBE) since 2004, and this has changed the direction of measurement and assessment in teaching and learning as ongoing quality improvement (Ahmad et al., 2011). The OBE methodology has to be continuously evaluated, quantified, and tracked using a variety of assessment methods through the mapping of Course Learning Outcomes (CLO). OBE may assess a student's total competency, which can be shown in a variety of abilities including teamwork and communication (presentations and reports), synthesis (case study analysis), practical skills (workshops and labs), etc. The CLOs, which are broken down into many carefully chosen evaluation techniques, clearly outline these talents. After that, the evaluation is created using the cognitive, psychomotor, and affective generic abilities from Bloom's Taxonomy and is given a grade. Learning outcomes are statements that specify what students should be able to do or demonstrate by the end of the learning process. They are distinct from learning goals and are directly linked to students, providing clear guidance on what they need to achieve throughout a course or program (Nasralla et al., 2021).

* Assoc. Prof. Dr., Bartın University, Faculty of Education, Bartın, Türkiye, volkanyuzuak@bartin.edu.tr, ORCID ID: 0000-0002-4712-0259
**Research Assist., Bartın University, Faculty of Education, Bartın, Türkiye zocal@bartin.edu.tr , ORCID ID: 0009-0002-2722-1199

To cite this article:
Yüzüak, A. V., & Öcal, Z. M. (2025). Assessment of student performance for laboratory application course examination using Rasch Measurement Model. *Journal of Measurement and Evaluation in Education and Psychology, 16*(2), 59-70. https://doi.org/10.21031/epod.1562806

Received: 7.10.2024
Accepted: 19.03.2025

Outcome-Based Education (OBE) (Rao, 2020; Spady, 1994) has recently gained significant attention in higher education, focusing on achieving specific targets and measurable results. Curriculum revisions are conducted periodically based on the exit learning outcomes that students must demonstrate upon completing a course or program. Learning outcomes encompass essential knowledge, skills, and values and are categorized into two main levels:

(a) **Program Learning Outcomes (PLOs)** – These define what graduates of a program should be capable of achieving.

(b) **Course Learning Outcomes (CLOs)** – These describe the skills and abilities students should demonstrate upon completing a course.

There is a strong connection between course-level learning outcomes and the core competencies of a program. Key Performance Indicators (KPIs) set at the program level can also be assessed at the course level. While course objectives outline the content a faculty member will cover, CLOs are student-centered and focus on what students will accomplish. A CLO statement consists of three key elements: an action verb indicating the required performance, a learning statement describing the acquired knowledge or skill, and a criterion defining the standard for acceptable performance (Hadj-Kacem et al., 2022).

Traditional educational methods that just evaluate students based on their final grade frequently fall short in measuring additional abilities that complement the content. However, CLOs measurement may give professors information about the knowledge and ability levels of their students, which can help them choose the best teaching and learning strategies for the course. The degree to which students can use the necessary abilities, and the efficacy of the instruction may both be determined by their achievement of the CLOs (Ahmad et al., 2011).

The evaluation of students' success has been based on their overall performance in completing assignments, quizzes, midterms, projects, and final examinations. Academics can utilize evaluation and measurement of the general performance output to guide them in establishing the best way to develop their teaching methodology and the caliber of their prepared questions. This provides an example of how the CLO for each course is fulfilled. Therefore, it is essential to use a reliable measuring technique to gauge and forecast students' performance in the future. This will also enable us to identify the kids who are most likely to struggle. The function of evaluation is evolving. As evaluation is now seen as a tool for learning rather than a way to track it, it is for learning. To determine where students are in their learning, where they need to go, and the best way to get there, instructors and students must search for and analyze data. This process is known as assessment for learning (Talib, Alomary & Alwadi, 2018). Assessment should not just be considered as something separate from teaching, delivered at the conclusion of the learning process, but also as a potent instrument for encouraging deep learning activities. Every evaluation, including those that are created and carried out by outside experts, those that are set by instructors at the conclusion of a certain work phase, and those that arise naturally in everyday classroom interactions between teachers and students, is significant. It is necessary to discuss the emergence of assessment cultures considering contemporary perspectives on learning, education, and the social role of assessment (Birenbaum & Dochy, 2009). Assessment culture is defined as educational evaluation practices that are consistent with prevailing ideologies, social expectations, attitudes, and values. In order to examine and validate the psychometric part of measuring instruments, the Rasch measurement model has been used extensively in modern times (Musa et al. 2011; Hamzah et al. 2011; Osman et al. 2012; Osman et al. 2013). The Rasch Model serves as an appropriate measurement method, offering a comprehensive and detailed understanding of individual abilities while also highlighting the features of each evaluation item (Manggopa & Batmetan, 2024).

It has been discovered that the Rasch model is appropriate for examining and validating people' (students') ability in relation to their course grades, as well as item difficulty (CLOs). The Rasch model is a different, "modern" approach to measuring that offers a reliable platform that satisfies the requirements of the SI Unit, acting as a repeatable instrument with a specified unit (Hamzah et al., 2011). This measurement model converts empirical data into logit scales with equal intervals by using data straight from the lecturer's assessment of a student for a particular job (Osman et al. 2013). Rasch converts the evaluation findings into a linear correlation by using logit as the measuring unit. Rasch shifts the focus of dependability from finding the data's "best fit line" to creating a dependable, repeatable measuring tool. Instead of fitting the data to the measurement model, it concentrates on building the measurement apparatus (Saidfudin et al., 2010).

Communities that support the tenets of learning cultures acknowledge the complexity of intelligence and work to provide all kids with the chance to learn in ways that are appropriate for their linguistic, cultural, and social backgrounds without pre-identifying their projected abilities. Birenbaum & Dochy (2009) asserts that the evaluation culture aligns with the constructivist pedagogy. According to this method, learning is seen as a process in which the student generates meaning; the instructor takes on the role of a mentor, giving students the chance to apply the information and abilities they already possess to comprehend new material. It is required of the instructor to assign engaging and difficult assignments. However, the factors contributing to the disparity in outcomes include the variations in learning styles that come from a range of backgrounds and cultures. Variations in assessment preferences are associated with variations in learning approaches, according to a 2006 study by David Gijbel using 108 first-year bachelor's students as a sample. In contrast to surface methods to learning, which focus on memorization and replication of the study materials' factual contents, deep approaches to learning are linked to students' goals to comprehend and develop the meaning of the acquired knowledge (Utaberta & Hassanpour, 2011). One way to assess students is by assigning assignments, examinations, and quizzes at different points during the 14-week study session each semester. Based on the students' performance, which indicates their level of learning, CLOs were assessed. The marks obtained are on a continuum scale, even though they are ordered. Therefore, it is relatively shallow and unfeasible to conduct additional examination utilizing the Traditional Test Method's raw score for this purpose (Mohamed et al. 2008). This research presents an alternate method that measures the CLO more accurately by utilizing the bio-based Rasch Unidimensional Measurement Model (WinSteps) as a performance evaluation tool. WinSteps converts the evaluation findings into a linear correlation by using "logit" as the measurement unit (Zoubir & Iskander, 2007). A synopsis of the measuring model is provided, along with illustrations of its main idea. A decision support system (DSS) course evaluation was created with each dimension of the skills to be measured shown using Bloom's Taxonomy, a behavioral science learning characteristic. The results as mentioned above were examined based on their degree of correlation with the qualities under examination (Thompson, 2007). It is also cross-checked for compatibility with the CLO's maps and serves as a roadmap for future iterations of the instructional approach and style. It gives instructors a more precise and comprehensive understanding of the degree of learning competency attained by students (Mohamed, et al. 2008).

According to Rozeha et al. (2008), the Rasch Model is a novel measuring technique that converts assessment data from students into a "logit" scale, which results in a linear correlation with an equal interval for the assessment outcome. As per Azrilah et al. (2008), in Rasch, the outcome was a dependable and consistent measuring tool rather than the determination of the "best fit line." The outcomes were then examined to see if they had been appropriately assessed, and the lecturer would thereafter utilize them as direction to enhance the way they teach (Rozeha et al., 2008). Because Rasch focuses on building the measuring instrument with precision rather than tailoring the data to suit a measurement model with mistakes, the findings of his research will give lecturers more accurate data on the learning capacity success of their students (Azrilah et al., 2008). Compared to more conventional approaches, the Rasch Model is anticipated to provide an accurate assessment of the CLO's performance.

Therefore, in this research, a measuring technique based on the Rasch Model is presented. This approach has been applied in the past to market research, psychometrics, and the health sector in addition to the educational area. It may be used to evaluate data for performance evaluation, skill measurement, or performance assessment. By assessing CLO using the Rasch Model, it is hoped to improve both the learning process and the outcome in comparison to other conventional testing techniques that don't offer a reliable mechanism with accurate results or a consistent way to evaluate CLO based on students' perspectives (Ahmad et al., 2010). In order to determine the students' competency in each CLO, the purpose of this study is to create a Rasch assessment scale. The same scale may be used to tabulate the analysis of students' competency and CLO difficulty. Proper improvement activities may be designed for the weaker students based on the findings, readying them for the higher courses. Therefore, the data can be used as a benchmark by the instructors to further enhance the method of teaching and learning for foundational courses in this program. The results can also be used to improve the way instruction is given and to give instructors ideas for assessing each CLO's performance. Therefore, in this study, the following question tried to be answered.

- How does the research present the students’ achievement on CLOs in Science Education course using Rasch Measurement Model?

Methods

Measuring CLO using Rasch Model

Within item response theory (IRT), the Rasch Model is a one-parameter logistic and static model that allows the amount of a latent trait present in an individual and the amount of the same latent trait reflected in different items to be estimated independently while still being explicitly compared to one another (Bradley et al., 2010). Each person with a certain level of a latent characteristic uses the Rasch Model to specify the likelihood that an answer will fall into one of the item's categories. Learning Performance Measurement System (LPMS) at Malaysian Institution of Higher Learning (IHL) now has the chance to examine the caliber of learning performance thanks to Rasch measurement for CLO evaluation. The following is the Rasch Model, which is used to demonstrate aptitude for a certain task:

$\Pr\{x_i = 1\} = e^{(\beta v - \delta i)} / [1 + e^{(\beta v - \delta i)}]$ (1)
where
e = Euler’s number, taken as 2.71828.
β_v = the ability of person v;
δi= the difficulty of assessment item i

By adding the log function, the preceding equation may be made even simpler to calculate the chance of success. The difference between ability measurement and item difficulty predicts the likelihood of success, or logit, as follows:

Logit (P/1-P) = $\beta v - \delta i$ (2)

Thus, the probability for a CLO achievement can be summarized as below:

Probability of success for a CLO = Ability of a student - Difficulty of a given task (3)

The likelihood of success for every CLO demonstrates how an individual's talent relates to how challenging the CLOs are.

Participants

A group of 18 students who enrolled in Science Education course during the spring semester of 2023-2024 was selected in the study. The course in Science Education was chosen as it is the must course for the science education students in this program and with the help of this course students can improve their laboratory abilities as prospective science teacher.

Data Collection Process

Planning Phase

This phase begins with defining the research domain. The Laboratory Application in Science Education course was selected as the research focus, and its course learning outcomes (CLOs) were analyzed. In essence, the course aims to equip learners with practical laboratory skills in science education. The development of the CLOs aligns with Bloom’s Taxonomy, encompassing cognitive learning levels such as knowledge, comprehension, application, analysis, evaluation, and synthesis, as outlined in Table 1. Various assessment methods were implemented in this course to evaluate students' understanding of the material. These assessments include lab reports (10%), presentations (25%), a midterm exam (40%), and a final exam (25%).

Table 1
CLO for Science Education Course

	Course Learning Outcomes	Bloom Taxonomy
CLO1	Understands and applies safety precautions in science laboratories.	Knowledge (C1)
CLO2	Recognizes laboratory materials and uses them for their intended purposes.	Comprehension (C2)
CLO3	Can use computer-aided laboratory applications.	Application (C3)
CLO4	Identifies and performs experiments with simple and cheap materials.	Application (C3)
CLO5	Makes explanations based on observations about the causes of events that have occurred.	Analysis (C5)
CLO6	Designs experiments by using science education curriculum and performs these experiments in a laboratory environment.	Synthesis (C6)
CLO7	Interprets the data obtained as a result of the experiment and the model created to reach relationships between concepts.	Evaluation (C7)

Classification Phase

All the questions selected for midterm and final exam; all the criteria for the lab report assignments; all the rubric items to evaluate the teaching presentations were chosen and classified according to CLO level. According to the classification, the distribution percentage of each question based on the CLO was summarized as indicated in Table 2.

Table 2
Percentage Distribution According to CLO Indices

	Lab-Report (10%)	Presentation (25%)	Midterm (40%)	Final Exam (25%)	100%
CLO1	0.20	0	0.25	0.05	13.25
CLO2	0	0	0.10	0	4.00
CLO3	0	0.18	0	0	4.50
CLO4	0.10	0	0.40	0	17.00
CLO5	0.35	0	0.05	0.30	13.00
CLO6	0.05	0.70	0.10	0.05	23.25
CLO7	0.30	0.12	0.10	0.60	25.00
Check	1.00	1.00	1.00	1.00	100

When Table 2 is examined, it can be understood that CLO6 and CLO7 have the majority of the percentage among them. These were decided depending on the experience of the course instructor.

Data Analysis

Subsequently, the CLO marks percentage distributions were recorded. To get the total for each CLO, the assessment marks for each were added up and divided by the total. A breakdown of student marks according to CLO is provided in Table 3. Additionally, it was evaluated that based on students' gender (male = M, female = F), in an effort evaluated to identify any differences in academic performance along these lines.

Table 3

Marks Distribution According to CLO

		CLO Achievement						
Student	**Gender**	**CLO1**	**CLO2**	**CLO3**	**CLO4**	**CLO5**	**CLO6**	**CLO7**
STD 1	F	68.47	61.92	73.14	67.70	81.85	88.93	58.22
STD 2	M	55.80	57.93	70.26	73.40	58.12	85.33	53.74
STD 3	F	79.17	52.19	63.90	88.75	52.83	89.88	59.40
STD 4	F	68.97	64.11	53.68	57.45	72.55	76.61	74.06
STD 5	F	66.80	53.05	73.50	78.20	82.32	89.38	88.78
STD 6	F	34.77	47.54	69.15	62.45	76.67	83.94	52.01
STD 7	M	43.77	40.16	68.07	60.65	75.27	82.59	58.33
STD 8	F	67.57	31.59	62.17	66.35	57.59	77.72	56.71
STD 9	M	63.77	66.16	58.07	70.65	73.87	72.59	67.33
STD 10	F	81.47	33.05	56.38	72.20	76.05	86.98	64.40
STD 11	F	78.10	51.79	32.75	57.15	61.34	78.44	56.91
STD 12	F	76.17	68.19	63.90	88.75	62.83	73.88	59.40
STD 13	F	62.13	64.18	53.86	84.70	66.79	75.83	59.34
STD 14	F	52.13	39.25	66.30	68.20	42.49	70.38	57.58
STD 15	F	59.80	52.43	34.58	62.70	33.72	90.73	58.28
STD 16	F	59.87	51.45	74.66	89.80	83.81	87.82	69.34
STD 17	F	48.50	68.19	52.38	72.75	67.90	75.48	61.48
STD 18	F	37.10	41.61	43.67	67.85	69.94	69.79	55.93

Marks for each CLO were then assigned according to grade based on the category below.

$$F(x) = \begin{cases} 0 \text{ if } 0 \leq x < 40; \\ 1 \text{ if } 40 \leq x < 50; \\ 2 \text{ if } 50 \leq x < 60; \\ 3 \text{ if } 60 \leq x < 70; \\ 4 \text{ if } 70 \leq x < 80; \\ 5 \text{ other}; \end{cases}$$

Table 3 shows the CLO achievement of the students. It can be interpreted that while students are good at mostly analyzing and synthesizing the concepts related to the course, they have difficulty in the comprehension and the application level of Bloom's Taxonomy related to the phenomenon in the laboratory application in science education. Before the CLO marks were entered into the Winstep program, they were first mapped into a grade category. Table 4 displayed the mapping procedure results.

Table 4

Mapping Results

		CLO Achievement						
Student	Gender	CLO1	CLO2	CLO3	CLO4	CLO5	CLO6	CLO7
STD 1	F	3	3	4	3	5	5	2
STD 2	M	2	2	4	4	2	5	2
STD 3	F	4	2	3	5	2	5	2
STD 4	F	3	3	2	2	4	4	4
STD 5	F	3	2	4	4	5	5	5
STD 6	F	0	1	3	3	4	5	2
STD 7	M	1	1	3	3	4	5	2
STD 8	F	3	0	3	3	2	4	2
STD 9	M	3	3	2	4	4	4	3
STD 10	F	5	0	2	4	4	5	3
STD 11	F	4	2	0	2	3	4	2
STD 12	F	4	3	3	5	3	4	2
STD 13	F	3	3	2	5	3	4	2
STD 14	F	2	1	3	3	1	4	2
STD 15	F	2	2	0	3	0	5	2
STD 16	F	1	2	4	5	5	5	3
STD 17	F	1	3	2	4	3	4	3
STD 18	F	0	1	1	3	3	3	2

On the other hand, while the sample size is relatively small, conducting a DIF (Differential Item Functioning) analysis based on gender could reveal interesting insights. In order to provide richer conclusions of potential subgroup differences and understand whether or not they represent DIF based on gender and CLOs was examined in the study. Table 5 indicates the DIF results based on the variable of gender.

Table 5

DIF Results Based On The Variable Of Gender

Gender	**Obs.-Exp Avrg.**	**DIF Measure**	**DIF S.E.**	**Gender**	**Obs.-Exp Avrg.**	**DIF Measure**	**DIF S.E.**	**DIF contrast**	**Joint S.E.**	**Rasch-Welch t**	**df**	**Prob.**	**Mantel chi-squ**	**Prob.**	**Active slices**	**CLO number**	**Name**
F	.10	.07	.27	M	-.51	.71	.57	-.63	.63	-1.00	2	.4217	1.5000	.2207	1	1	CLO1
F	-.01	.75	.26	M	.05	.71	.57	.05	.63	.08	2	.9458	1.1429	.2850	1	2	CLO2
F	-.09	.21	.26	M	.44	-.36	.62	.58	.67	.86	2	.4794	.5000	.4795	1	3	CLO3
F	.01	-1.16	.30	M	.00	-1.16	.65	.00	.71	.00	2	1.0000	.5000	.4795	1	4	CLO4
F	-.02	-.59	.28	M	.11	-.75	.63	.16	.69	.23	2	.8395	.5000	.4795	1	5	CLO5
F	-.03	-2.46	.38	M	.15	-2.95	1.05	.49	1.12	.44	2	.7031	2.0000	.1573	1	6	CLO6
F	.05	.07	.27	M	-.23	.37	.59	-.30	.65	-.46	2	.6932	.5000	.4795	1	7	CLO7
M	-.51	.71	.57	F	.10	.07	.27	.63	.63	1.00	2	.4217	1.5000	.2207	1	1	CLO1
M	.05	.71	.57	F	-.01	.75	.26	-.05	.63	-.08	2	.9458	1.1429	.2850	1	2	CLO2
M	.44	-.36	.62	F	-.09	.21	.26	-.58	.67	-.86	2	.4794	.5000	.4795	1	3	CLO3
M	.00	-1.16	.65	F	.01	-1.16	.30	.00	.71	.00	2	1.0000	.5000	.4795	1	4	CLO4
M	.11	-.75	.63	F	-.02	-.59	.28	.16	.69	-.23	2	.8395	.5000	.4795	1	5	CLO5
M	.15	-2.95	1.05	F	-.03	-2.46	.38	.49	1.12	-.44	2	.7031	2.0000	.1573	1	6	CLO6
M	-.23	.37	.59	F	.05	.07	.27	-.30	.65	.46	2	.6932	.5000	.4795	1	7	CLO7

Width of Mantel slice: MHSLICE = .010 logits, Zero cell adjustment: MHZERO = .0000

According to the results of Table 5, there is no significant difference in p-values for all CLOs. The DIF Contrast or Mantel-Haenszel test does not reveal a significant difference for any gain (CLO1-CLO7). This indicates that the expressed gains offer a fair assessment of the gender variable. It means that the CLOs work the same way for male and female students.

Results and Discussion

WinSteps 3.69 software is used to perform the tabulation of assessment results for a group of eighteen students and compute the associated result. The Person-Item Distribution Map (PIDM) is produced by the program following the processing of the input data. Each student's placement in relation to the CLOs distribution (Item) is displayed in Figure 1 (Person=STDnn GenderX). In accordance with the Latent Trait Theory, the PIDM plots the distribution of Person and Item on the same logit scale. The PIDM shows a person's ability β in reaction to an item's difficulty δi. According to Rashid and Zaharim (2007), the parameter β represents the item's placement on the same characteristic. If βn is bigger than δi, the individual is more likely to be able to reply to the item properly.

The degree of an individual's competence is represented by the distance between the item and the individual's position on the map; the greater the distance, the greater the likelihood that the individual would correctly answer the given item (Rashid & Zaharim, 2007). In the meanwhile, an item's difficulty is determined by how widely it falls on a scale. For example, an item located farther from the Meanitem is harder than an item positioned closer. The Meanitem, which acts as the logit scale threshold, is set to zero in this study.

Figure 1

Person-Item Distribution Map

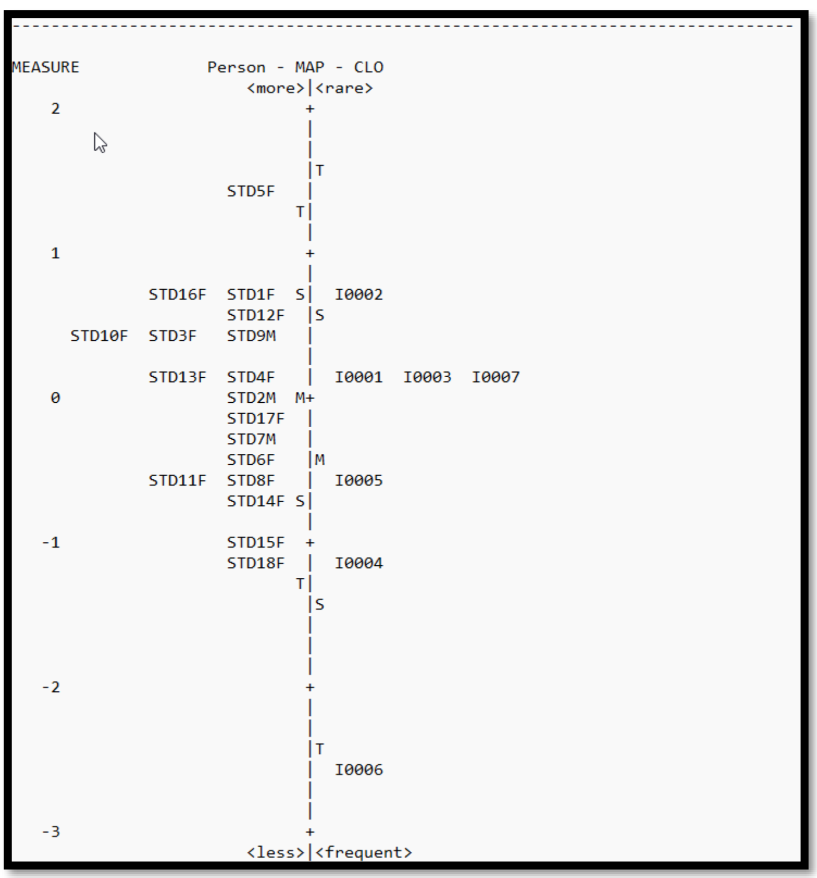

Figure 1 represents the density of the students for the CLOs depending on the achievement locations. The Meanitem acts as a benchmark and is assigned a value of zero on the logit scale. Items with higher Meanitem locations are considered more challenging compared to those at lower positions. Similarly, in the distribution of individuals, high-performing students are positioned at the upper end of the map, while lower-performing students are represented at both extremes. As a result, an individual's ability level can be assessed using person-item distribution map (PIDM) by examining the gap between individuals and items on the map. A larger separation indicates a higher likelihood of successfully achieving the item (Bradley et al., 2010).

To assess each student's and CLO's performance in the PIDM, the logit values are produced as indicated in Tables 6 and 7. The logit value position for each student and CLO is displayed via the STD and CLO measures. The cohort's Meanperson = -0.34 is less than the threshold value, Meanitem = 0.00, according to the PIDM. This suggests that students' proficiency on the assigned CLOs is low. It was determined that eight of the eighteen students (44.4%) did not meet the Meanitem. The remaining CLOs are challenging for these students to meet, even though they are able to meet CLOs 4, 5, and 6. The item with the highest difficulty level among the seven CLOs was CLO2 (logit 0.75). On midterm and final exams, the majority of the questions used to assess CLO1 and CLO2 are found. In order for students to answer the questions on the midterm and final test, they must commit the information to memory. When it comes to answering questions without consulting books or notes, this might be challenging for students. The high logit values of CLO2 might perhaps be attributed to this. Table 6 shows logit values for each student and CLO.

Table 6

Logit Value for Each Student and CLO

Student	Logit value	Student	Logit value	CLO	Logit value
STD5F	1.39	**STD2M**	0.03	**CLO1**	0.18
STD1F	0.76	**STD17F**	-0.13	**CLO2**	0.75
STD16F	0.76	**STD7M**	-0.29	**CLO3**	0.12
STD12F	0.57	**STD6F**	-0.45	**CLO4**	-1.16
STD3F	0.39	**STD8F**	-0.60	**CLO5**	-0.62
STD9M	0.39	**STD11F**	-0.60	**CLO6**	-2.52
STD10F	0.39	**STD14F**	-0.76	**CLO7**	0.12
STD4F	0.21	**STD15F**	-1.05		
STD13F	0.21	**STD18F**	-1.20		

The majority of the student distribution regarding CLO is concentrated over the threshold value overall. For example, among all the students, student coded as STD5F is ranked top. With a logit score of 1.39, he does well overall compared to the projected CLO performance. In order to only be able to complete CLO (Synthesis), the lowest-performing student, STD18F (logit -1.2), needs to work harder. Differential Item Function (DIF) is a feature offered by Winstep program. This feature allows the PIDM to distinguish between gender differences in learning capacity.

With a logit value of -2.52, CLO6 (synthesis) is the most easily exposed item in the PIDM. Depending on the presentations, CLO6 is assessed using a performance-based evaluation methodology. According to Bloom's Taxonomy rating, this item should be the hardest to complete, however PIDM shows that all students understand it well. This is due to the fact that CLO6 grades are determined by evaluating not just the paperwork but also the students' cooperation, attitude, and presentation-related efforts. The likelihood that each student will meet each CLO is presented in Table 7. By calculating the likelihood that each student will attain the CLOs, it provides a detailed examination of the link between each individual and each item. Equations (1) and (2) can also be used to compute it manually.

When computing the likelihood of obtaining CLO2, using student STD5F as an example, equation (2) yields the value of P(θ), which is as follows:

P(θ) = βv (STD5F) – δi (CLO2)
= 1.39 – 0.75
= 0.64
Substitute this value into equation (1):
P(θ) = e^(βv - δi) / [1 + e^(βv - δi)]
=0.65

The value of 0.65 will be the CLO2 achievement of student STD5F. The rest of the analysis is given in Table 7.

Table 7
Probability of Each Student to Achieve Each CLO

Probability of success	**Item**						
	CLO1	CLO2	CLO3	CLO4	CLO5	CLO6	CLO7
P(STD5F)	0.77	0.65	0.78	0.93	0.88	0.98	0.78
P(STD1F)	0.64	0.50	0.65	0.87	0.80	0.96	0.65
P(STD16F)	0.64	0.50	0.65	0.87	0.80	0.96	0.65
P(STD12F)	0.60	0.46	0.61	0.85	0.77	0.96	0.61
P(STD3F)	0.55	0.41	0.57	0.82	0.73	0.95	0.57
P(STD9M)	0.55	0.41	0.57	0.82	0.73	0.95	0.57
P(STD10F)	0.55	0.41	0.57	0.82	0.73	0.95	0.57
P(STD4F)	0.51	0.37	0.52	0.80	0.70	0.94	0.52
P(STD13F)	0.51	0.37	0.52	0.80	0.70	0.94	0.52
P(STD2M)	0.46	0.33	0.48	0.77	0.66	0.93	0.48
P(STD17F)	0.42	0.29	0.44	0.74	0.62	0.92	0.44
P(STD7M)	0.38	0.26	0.40	0.70	0.58	0.90	0.40
P(STD6F)	0.35	0.23	0.36	0.67	0.54	0.89	0.36
P(STD8F)	0.31	0.21	0.33	0.64	0.50	0.87	0.33
P(STD11F)	0.31	0.21	0.33	0.64	0.50	0.87	0.33
P(STD14F)	0.28	0.18	0.29	0.60	0.47	0.85	0.29
P(STD15F)	0.23	0.14	0.24	0.53	0.39	0.81	0.24
P(STD18F)	0.20	0.12	0.21	0.49	0.36	0.79	0.21

Table 7 indicates that just three students (16.6%) out of the total of 18 students do not have any difficulties achieving their CLOs. The remaining fifteen students are largely struggling to achieve CLO2, where it is indicated in bold that the likelihood of attaining CLO is less than 0.50. CLO6 (Synthesis) is the finest CLO accomplishment since it allows all students to do well depending on the likelihood displayed. According to the table, every student can attain a high CLO as opposed to a low CLO. This is a common occurrence as CLO6 (synthesis), unlike other CLOs, is only assessed during the presentation; in contrast, it is examined at the midterm, lab report, and final exam. The students are given plenty of time to organize, brainstorm, and finish their presentations. In contrast to midterm, lab reports, and final exams, the exercises are completed alone in a setting comparable to an exam.

Conclusion

This study has demonstrated that the Rasch Model provides more accurate results when used to assess CLO performance for Science Education course. When compared to conventional approaches, which just use the distribution of questionnaire forms to estimate the CLO based on students' assumptions, this type of measurement is superior. The association pattern between students and each CLO's performance level may be generated using this model. It is not possible to create this pattern with traditional measuring techniques. The study's findings can help lecturers track students' progress toward each of the course's specified CLOs. In addition to highlighting underachievers, CLO performance indicates how well the lecturer is instructing. This study concludes that the performance of science education students is generally satisfactory and can be effectively assessed using the Rasch measurement model. The results also demonstrate a pattern consistent with traditional evaluation methods. Through Rasch analysis, students are categorized based on their achievements in various learning tasks, highlighting their abilities in the course. The study reveals that the Rasch model can establish relationships between students and their performance levels for specific tasks, which conventional methods fail to achieve. This makes the Rasch model a superior tool for evaluating learning performance. Additionally, the insights from Rasch analysis can serve as a valuable resource for lecturers to track and monitor their students' progress. The study further establishes that the Rasch model is a reliable and effective instrument for performance evaluations, particularly in quality-focused systems, as it provides a more precise analysis of student performance, resulting in clearer and more comprehensible evaluations.

Besides all these, the efficacy of teaching and learning may be thoroughly and widely examined using this straightforward yet cautious conceptual theoretical framework of measurement. For a particular education method—in this example, active learning—Rasch measurement employs actual data straight from the student evaluation. The students were more precisely categorized based on their observed accomplishments when the Rasch Measurement was used. It allows for the separate evaluation of each component. The growth of upcoming "ingenious" graduates might greatly benefit from the continued use of Rasch Measurement in Science Education.

Declarations

Conflict of Interest: The authors reported no potential conflict of interest.

Ethical Approval: The study utilized course data. The authors confirm that all ethical guidelines were followed.

References

Ahmad, K., Abdullah, S. & Hood, Z. (2010). Kajian Keberkesanan Kaedah Pengajaran Dan Pembelajaran Kursus Analisis Dan Reka Bentuk Sistem. *Proceedings of Teaching and Learning Congress.*

Ahmad, K., Mohamad, N.A., & Zainudin, S. (2011). An improved course assessment measurement for analyzing learning outcomes performance using rasch model. *Procedia - Social and Behavioral Sciences*, *18*(1), pp. 442-449. https://doi.org/10.1016/j.sbspro.2011.05.065

Azrilah, A. A., Saidfudin, M., NoorHabibah A., Zaharim, A., Azlinah M., Ghulman, H.A. & Sohaimi, Z. (2008). Application of Rasch Model in validating the construct of measurement instrument. *International Journal of Education and Information Technologies*, *2*(2), pp. 105-112.

Birenbaum, M. & Dochy, F. (2009). Alternatives in Assessment of Achievement, Learning Processes and Prior Knowledge, (Eds). Norwell MA: Kluwer. https://doi.org/10.1007/978-94-011-0657-3

Bradley,K.D., Cunningham, J. , Haines, T., Mueller, C.E., Royal,K.D., Sampson, S.O., Singletary, G., & Weber, J.A. (2010). Constructing and Evaluating Measures: Applications of the Rasch Measurement Model. http://www.uky.edu/~kdbrad2/Rasch_Symposium.pdf.

Hadj-Kacem, Y., Alshehri, S., & Qaid, T. (2022). Categorizing well-written course learning outcomes using machine learning. *Journal of Information Technology Education: Innovations in Practice, 21*, 61-75. https://doi.org/10.28945/4997

Hamzah, N., Khoiry, M., Osman, S., Hamid, R., Jaafar, O. & Arshad, I. (2011). Poor student performance in project management course: was the exam wrongly set? *Recent Researches in Education,* pp. 154-160. https://doi.org/10.1109/ICEED.2017.8251175

Manggopa, H.K., & Batmetan, J.R. (2024). Evaluation of students' achievement in the development of innovative learning tools for ppg computer engineering and informatics students using the rasch model. *International Journal of Information Technology and Education (IJITE)*, *3*(2), 54-65. https://doi.org/10.62711/ijite.v3i2.177

Mohamed, A., Aziz, A., Ahlan, R., Zakaria, S., & Masodi, M. (2008). Outcome based education performance measurement: a rasch-based longitudinal assessment model to measure information management courses LO's. *Wseas Transactions On Information Science & Applications*, *3*(5), 292-299.

Musa, N.C., Mahmud, Z. & Ghani, N.M. (2011). Application of rasch measurement model in validation and analysis of measurement instruments in statistical education research. *Recent Researches in Education*, pp. 123-129. http://www.wseas.us/e-library/conferences/2011/Penang/EDU/EDU-19.pdf

Nasralla, M.M., Al-Shattarat, B., Almakhles, D.J., Abdelhadi, A., & Abowardah, E.S. (2021). Futuristic trends and innovations for examining the performance of course learning outcomes using the rasch analytical model. *Electronics, 10*, 727. https://doi.org/10.3390/electronics10060727

Osman, S., Khoiry, M., Badaruzzaman, W. & Mutalib, A. (2013). "Measurement Of Students' Understanding In Final Examination Of Statics And Dynamics Course Using Rasch Measurement Model", IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE). https://doi.org/10.1109/TALE.2013.6654549

Osman, S., Naam, S., Jaafar, O., Badaruzzaman, W. & Rahmat, R. (2012). Application of rasch model in measuring students' performance in civil engineering design II course. *Procedia-Social and Behavioral Sciences*, *56*, pp. 59-66.https://doi.org/10.1016/j.sbspro.2012.09.632

Rao, N. J. (2020). Outcome-based education: An outline. *Higher Education for the Future, 7*(1), 5-21. https://doi.org/10.1177/2347631119886418

Rashid, A. & Zaharim, A., M. (2007). Application of Rasch Measurement in Evaluation of Learning Outcomes: A Case Study in Electrical Engineering. Regional Conference on Engineering Mathematics, Mechanics, Manufacturing & Architecture (EM3ARC), UKM, Bangi. https://doi.org/10.1016/j.sbspro.2012.09.632

Rozeha, A.R., Razimah, A., Hamzah, A.G., & Saidfudin, M. (2008). Application of Rasch-based ESPEGS model in measuring generic skills of engineering students: A New Paradigm. *WSEAS Transactions on Advances in Engineering Education, 5*(8), 591-602.

Saidfudin, M., Azrilah, A. A., Rodzo'an, N. A., Omar, M.Z., Zaharim, A. & Basri, H. (2010). "Easier Learning Outcomes Analysis using Rasch Model in Engineering Education Research," EDUCATION'10 Proceedings of the 7th WSEAS International Conference on Engineering Education, pp. 442-447. https://dl.acm.org/doi/abs/10.5555/1864181.1864264

Spady, W. G. (1994). Outcome-based education: Critical issues and answers. American Association of School Administrators. https://files.eric.ed.gov/fulltext/ED380910.pdf

Talib, A.M., Alomary, F.O. & Alwadi, H.F. (2018). Assessment of student performance for course examination using rasch measurement model: A case study of information technology fundamentals course. *Hindawi Education Research International*, *2018*(1), 1-8. https://doi.org/10.1155/2018/8719012

Thompson, D., (2007). Jacknifing Multitaper Spectrum Estimates. IEEE Signal Processing Magazine, 13-20. https://doi.org/10.1109/MSP.2007.4286561

Utaberta, N. & Hassanpour, B. (2011). Aligning assessment with learning outcomes. UKM Teaching and Learning Congress 2011. doi: 10.1016/j.sbspro.2012.09.372

Zoubir, A.M., & Iskander, D.R. (2007). Bootstrap methods and Applications in Signal Processing. *IEEE Signal Processing Magazine*, 10-19.