JOURNAL OF BALTIC SCIENCE EDUCATION

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

A MANY-FACET RASCH MEASUREMENT APPROACH TO ANALYZE THE PREPARED SCIENCE LABORATORY ACTIVITIES BASED ON SCIENCE PROCESS SKILLS AND VIEWS OF PRE-SERVICE SCIENCE TEACHERS

Emrah Hiğde,
Ahmet Volkan Yüzüak,
Zekiye Merve Öcal,
Hilal Aktamış

Abstract. *The Many-Facet Rasch model is frequently used to analyse and minimize disparities in rater (judge) severity in performance evaluations, in which raters assign scores to test-takers' performances. In this research, the aim of the present study was to analyse science teacher candidates' laboratory activities by using the Many-facet Rasch model. Rasch model's surfaces are, respectively: 9 juries, 8 science activities and 24 criteria. The FACETS program was used to do data analysis. Findings show that laboratory activities, which were coded as E8, were found to be the most successful and E2coded activity was found to be the least successful based on the criteria. Jury numbered 7 or coded as G7, is the most lenient, and scorer numbered 5, or coded as G5, is the severest when the juries are listed from the most lenient. The study's objective is to use the Many-Facet Rasch measurement model to analyse laboratory experiments linked to science-related activities. Analysis of the performance of science activities, analysis of criteria hardness, analysis of the severity and leniency of juries, and study of jury bias were carried out concurrently with this goal. At the end of the study, it can be easily inferred that the Multi-Facet Rasch measurement model could be used effectively to evaluate peer groups in science education and objective results could be obtained.*

Keywords: *laboratory experiments, Rasch model, science activities, science education*

Emrah Hiğde
Aydın Adnan Menderes University, Türkiye
Ahmet Volkan Yüzüak
Bartın University, Türkiye
Zekiye Merve Öcal
Bartın University, Türkiye
Hilal Aktamış
Aydın Adnan Menderes University, Türkiye

Introduction

Science education aims to ensure that every individual participating in society has various basic and complex cognitive characteristics, consciousness, attitudes and values. In this regard, ensuring that individuals participate in society as individuals who can question, research, make inferences based on scientific foundations and make decisions in line with these inferences, and use, apply and develop science and technology is related to the quality of education in schools (Abd-El Khalick, et al., 1998). The trained workforce that countries need can only be provided by individuals graduating from the education system having the necessary skills. The development levels, economic situations and social welfare levels of countries are directly related to the skills of individuals participating in the workforce (Organization for Economic Co-operation and Development [OECD], 2019). All individuals need to acquire science process skills to use a rational perspective in their decisions, benefit from data and facts, and evaluate events and situations from a scientific perspective. In addition, having science process skills makes it easier for individuals to turn to science, technology and science-related professions, providing a high-value-added workforce to the country's economy (Hamarat & Arkan, 2018; OECD, 2019). For this reason, measuring the science process skills of individuals, and completing and improving them can be seen as important for both the individual and the country's future. Therefore, scientific process skills could be acquired by individuals during their educational lives. To achieve this, teachers should understand the nature of science to know how to reach scientific knowledge, just like a scientist. Because the teacher interprets and transfers his/her own knowledge and experience to the student during the lesson (Hofstein & Lunetta, 1982).

However, examining the level at which these skills are acquired is the most important step in the educational process. In order to ascertain the degree of science process skills attained by science teacher candidates, it

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

was crucial to employ one of the most significant phases of the measurement and evaluation teaching process in science education. Measuring science process skills, which had an important place in science education, would help to reveal potential deficiencies or strengths in pre-service teachers so that the necessary feedback correction studies could be carried out and future planning of teacher candidates could be made more accurately. When the studies containing all these classifications were examined, it was seen that pre-service teachers could acquire all skills through certain science activities (Harlen, 1999; Huppert, et al., 2002). Therefore, laboratory activities could help pre-service teachers learn scientific processes and give opportunities for the development of these skills by using these processes effectively (Duru et al., 2011). In particular, science teachers need to be familiar with scientific language and understand scientific processes in depth in order to teach science (Cotabish et al., 2011). Numerous investigations, however, have revealed that teachers lacked a theoretical understanding of these skills (Türkmen & Kandemir, 2011) and that science teacher candidates were not competent in teaching science process skills (Yılmaz-Tüzün & Özgelen, 2012). The majority of science teachers did not use skills such as identifying materials, observing, measuring and recording, collecting and interpreting experimental data in their lessons (Yandila & Komane, 2004). It showed that science teacher candidates' knowledge level about science process skills was insufficient (Farsakoğlu, et al., 2008). These findings demonstrated the necessity for teacher candidates—the future educators—to develop their science process skills. Three arguments were presented by Rowland et al. 1987 supporting the need for scientific education to use a laboratory approach centred on science process skills. The process approach, to start, placed a strong emphasis on science as a means of comprehending the natural world. Secondly, this approach would allow teacher candidates to better understand science subjects and the path followed by a scientist in the process of discovering events, principles and relationships in natural life. Thirdly, this approach would contribute to the development of scientific attitudes as it ensured the active participation of teacher candidates.

The scientific curriculum includes science process skills. The skills that scientists employed in their research included observing, measuring, categorizing, collecting data, formulating hypotheses, utilizing data to build models, modifying and controlling variables, and doing experiments (MoNE, 2018, p. 9). There was no clear approach to measuring and evaluating skills in the curriculum. A concentration on cognitive measurement of achievements rather than the measurement of science process skills in classroom assessment and evaluation processes resulted from the achievements' obvious lack of science process skills (Duruk et al., 2017; Wellington, 1989; Wu, 1994). Although measuring achievements was important in evaluating the quality of education, not measuring the science process skills contained in the achievements in detail caused the skill deficiencies of teacher candidates to remain hidden (Anderson & Krathwohl, 2001). As a result, it became difficult for professors, parents or institutions to carry out the necessary follow-up, evaluation and development activities regarding the science process skill levels of teacher candidates, even in general. Science process skills were processes that were expected to be used and put to work regardless of the classification method. Upon reviewing the literature, it was found that many researchers have limited their assessment of science process skills to multiple choice questions (Aydoğdu, 2017; Bahşı & Açıkgül Fırat, 2020; Ergül et al., 2011; Fathonah et al., 2018; Gill, 2019; Özgelen, 2012; Şensoy & Yıldırım, 2017; Uysal & Cebesoy, 2022). Given the nature of science process abilities, information gathered through multiple choice questions alone would not be adequate. Nevertheless, other research (Aktamış & Ergin, 2007; Aktamış & Şahin Pekmez, 2011; Azizah et al., 2018; Indri et al., 2020; Rillero, 1998; Serevina et al., 2018; Strong, 2013) also benefited from using open-ended questions and activities. When the research's limitations were taken into account, the primary issues were that all of these studies had a small sample size because they were conducted manually and face-to-face with participants (Leat & Nichols, 2000). Additionally, the study was not sustainable and there was insufficient data to comment on how the participants' science process skills changed over time (Bearman et al., 2020; Lederman & Stefanich, 2007; Paine, 2022; Richardson & Clesham, 2021; Webb, 2007). To address these issues, it would be suitable to use peer and self-evaluation within the parameters of the study and to make use of the Many-Facet Rasch Model to guarantee objectivity in the measurement and evaluation process.

Currently, a pivotal inquiry is: what methods exist for the objective evaluation of Laboratory Experiments grounded in Science Process Skills? This inquiry is central to the present research. A potential solution is found in the use of the Many-Facet Rasch Model (MFRM), which is rooted in Item Response Theory, as indicated by Semerci's series of studies from 2011 to 2012 and further supported by Yüzüak, Erten, Kara, and Kaptan's research between 2015 and 2019. MFRM yields two robust metrics: the separation and reliability indices. The reliability index can be equated to Cronbach's Alpha or KR-20, signifying the proportion of 'True Variance' in relation to 'Observed Variance'. High reliability scores, approaching 1.0, are preferred for both individuals and items, as highlighted by Linacre in 2010. An expansion of Rasch measurement models is the Many Facet Rasch Model (MFRM) (Rasch, 1980; Wright & Stone, 1979). The Many-facet Rasch (Fk) equation looks like this:

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

$$\log\frac{P_{nijk}}{P_{nijk-1}} = B_n - D_i - C_j - F_k$$

The equation reads as follows: Bn is the examinee's ability; Di is the item's difficulty; Cj is the severity of judge j; Fk is the additional difficulty overcome in being observed at the level of category k, relative to category k-1; and Pnijk is the probability of examinee n being awarded on item i by judge j a rating of k-1 (Linacre, 1989). Judges' ratings of examinee performances are often necessary for authentic measurement (Linacre, 1994). The secret to this is MFRM. The purpose of this study is to assess how well MFRM can be used to prepare laboratory experiments based on science process skills. The preparation of experiments based on scientific process skills does not include closed-ended questions. For this reason, the most important tool used in scoring high-level skills is the rubric. Raters make their grades according to the performance criteria in the scoring keys (Kan, 2007). However, the rater's decision is not based solely on performance. Different factors/surfaces affect scoring. These include difficulty level of the task/performance, rater strictness/generosity, ratee's past, etc. These factors, which are not included in the content of the measurement, may influence the validity of the scoring (Prieto & Nieto, 2014). In this context, when scoring scientific process skills, the scores are: Rater effect; strictness/generosity, halo effect, avoidance of outliers; The difficulty of selecting criteria for performance; Rating criteria (points on the rating scale have different meanings among raters) affect variables. The rater effect is not related to the performance of the individual being scored but is one of the characteristic features of the rater. This variable interferes with measurements as an error and threatens the validity of measurement results (Eckes, 2005).

Research Problem and Focus

With the multi-facet Rasch model (MRM), rater characteristics can be analysed as a facet in the measurement model (Köse, et al., 2016). The preparation of experiments based on scientific process skills does not include closed-ended questions. For this reason, the most important tool used in scoring high-level skills is the rubric. Raters make their grades according to the performance criteria in the scoring keys (Kan, 2007). However, the rater's decision is not based solely on performance. Different factors/surfaces affect scoring. These include difficulty level of the task/performance, rater strictness/generosity, ratee's past, etc. These elements could affect the validity of the score because they are not part of the measurement's substance (Prieto & Nieto, 2014). In this context, when scoring scientific process skills, the scores are: Rater effect; strictness/generosity, halo effect, avoidance of outliers; the difficulty of selecting criteria for performance; Rating criteria (points on the rating scale have different meanings among raters) affect variables. The rater effect is not related to the performance of the individual being scored but is one of the characteristic features of the rater. This variable interferes with measurements as an error and threatens the validity of measurement results (Eckes, 2005).

Research Aim and Research Questions

In this research, based on the multi-facet Rasch model (MRM), rater characteristics can be analysed as a facet of the measurement model. In parallel with this aim, not only the laboratory experiment performance, severity/leniency, criterion hardness of science teacher candidates and their bias but also the opinions of pre-service science teachers are obtained to be analysed.

The aim of this research was to use the Many-Facet Rasch measurement model to analyse laboratory experiments linked to science-related activities. For this purpose, performance analysis of science activities, criterion hardness analysis, jury severity/leniency analysis, and jury bias analysis were carried out. Throughout the study, the following research questions were tried to be answered:

1) How do the rater characteristics of pre-service teachers analyse the laboratory experiments depending on the science-related activities?
2) What are the opinions of the pre-service science teachers about using a Many-Facet Rasch measurement approach to analyse the prepared science laboratory activities based on science process skills?

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Research Methodology

General Background

The book "Some Probabilistic Models for Intelligence and Attainment Tests" by George Rasch (1960) has a description of the Rasch model, a one-parameter logistic model based on item response theory. The Rasch model and the Many-facet Rasch model approach have been applied in the following fields: language testing (Eckes, 2005); educational and psychological measurement (Ahmad, Ali & Zainudin, 2011; Chang & Engelhard, 2016; Çetin & Ihan, 2017; Ismail, et al., 2017; Kaya, et al., 2017; Köse, et al., 2017; Semerci, 2012; Semerci, 2011; Yılmaz & Sözer, 2018), and health sciences (Park, et al., 2018).

A related study was carried out with pre-service science teachers in the 2023-2024 academic year. In this study, the survey research strategy was applied. By the ethical research rules (Aydın Adnan Menderes University Protocol Number = E-84982664-050.04-506435, Decision Date: 29.02.2024 Meeting Number: 2024/2-XIX), the laboratory experiments were coded as E1, E2 … E8; criteria were coded as a qualitative problem, well-explained problem, well-content problem etc., and the student groups (jury) were coded as G1, G2, … G8; expert was coded as G9.

Participants

In this research, participants consist of 27 pre-service science teachers who are the 3rd-grade students of the Science Laboratory Applications Course in the Department of Science Teaching in a university located in the western region of Turkey. Convenience sampling was used for the research due to its ease of use for both sample and study execution. Pre-service science teachers first evaluated themselves by making self-evaluation and then other groups based on the rubric given to them, individually. Later, the faculty member who took the course made evaluations within the framework of these competencies. With the help of this model, the best component of scientific process skill, which group has the better scientific process skill than the others and the objectivity of the evaluation process have been tried to be revealed at the end of the study.

Instrument and Procedures

The quantitative data from the research were analysed by using the Many-facet Rasch model. The criteria included 24 items that refer to measuring the degree of ability and skills in laboratory applications in terms of considering and evaluating the efficacy of laboratory works as science teacher candidates. Criteria and abbreviations are indicated in Table 1 (Kaygısız et al., 2017). For validity and reliability issues, based on Item Response Theory, the Many-Facet Rasch Model (MFRM) presents a viable answer, as demonstrated by Semerci's 2011–2012 set of studies and further reinforced by Yüzüak, Erten, Kara, and Kaptan's 2015–2019 study. The separation and dependability indices are two strong indicators that MFRM produces. The dependability index, which represents the ratio of "True Variance" to "Observed Variance," is equivalent to Cronbach's Alpha, or KR-20. As Linacre (2010) pointed out, both persons and objects should have high dependability scores, close to 1.0. The Many Facet Rasch Model (MFRM) is an extension of Rasch measurement models (Rasch, 1980; Wright & Stone, 1979). This is how the many-facet Rasch (Fk) equation appears:

Table 1
Criteria and Related Abbreviations

C Nu	Criteria	Criteria Abbreviations
C1	The problem is worth for designing experiments to find a solution.	Qualitative problem
C2	The specified problem sentence is correct as a statement.	Well-explained problem
C3	The specified problem sentence is correct in terms of content knowledge.	Well-content problem
C4	The specified problem statement is solvable.	Solvable problem
C5	The hypothesis statement is aimed at solving the problem.	Hypothesis and problem are parallel

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

C Nu	Criteria	Criteria Abbreviations
C6	The hypothesis sentence is correct as a statement.	Well-explained hypothesis
C7	The dependent variable was determined correctly.	Correct dependent variable
C8	The independent variable was determined correctly.	Correct independent variable
C9	Control variables were determined correctly.	Correct controlled variable
C10	The design of the experiment is sufficient to solve the problem.	Design is for problem solving
C11	It was mentioned how to change the independent variable.	Changing independent variable
C12	It was mentioned how to measure the dependent variable.	Measuring dependent variable
C13	The materials in the experiment are suitable for the problem and design of the experiment.	Convenient materials
C14	The procedure steps are given in a logical integrity/sequence.	Step by step movement
C15	At least three tests were made to ensure the reliability of the experimental data.	Reliability tested (at least 3 times)
C16	Variables that needed to be controlled in the experiment were controlled.	Correct controlled variables
C17	Tables/graphs appropriate to the data related to the experiment were used.	Convenient tables/graphics
C18	Dependent and independent variable names are included in the table.	Tabled variables
C19	The units of the independent and dependent variables are written correctly.	Correct units of variables
C20	The measurements/findings in the experiment were calculated correctly.	Correct measurement calculation
C21	The data has been correctly placed in the table.	Correct tables with variables
C22	In the results, an explanation for the problem was made.	Results explain the problem
C23	As a result of the experiment, correct explanation was made in terms of content knowledge.	Well-explained content results
C24	The hypothesis was taken into account when evaluating the experimental results.	Hypothesis based evaluation

Data Analysis

Within the context of the Science Laboratory Applications-II course, the research group's science teacher candidates were theoretically instructed on the content and components of science process skills. Some activities were presented to teacher candidates as an example. In the course implementation, closed-ended, semi-open-ended and open-ended experiment activities were carried out with the teacher candidates throughout the one semester. In addition to these, open-ended experiments were implemented to the science teacher candidates within the scope of the Science Laboratory Application-II course. They were given a new scenario for each week, and they had to come to class by designing their experiments until the next class day. They have conducted their experiments during the class time, and they have taken notes of their data. All tasks were completed by science teacher candidates in groups, and they prepared an experiment report for every lab activity. After that, they have to come to class with their experiment reports as prepared for the following week.

In the course, firstly, the reports of each experiment carried out the previous week were evaluated together with all groups in the classroom and the course instructor, based on the rating scales, which were shared with the teacher candidates. For the following week, they have been conducting the next experiment and gathering the data. At the end of the study plan, the analysis of the data obtained was made with the FACETS 3.71.4 program, developed by Linacre (1993) and based on the Rasch measurement model.

Research Results

The three facets are laboratory experiments, criteria, and student groups as a jury. Table 2 represents the related data as the calibration map.

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

Table 2
Calibration Map Related to Laboratory Experiments

Measr	+Laboratory experiments	+Criteria	+Group	RATING
2 +				+ (5)
	E8			
	E7			
1 +				+
	E3	C13 C15		
	E5			---
		C14		
		C10 C6		4
	E6		G4 G5	
		C20 C5	G8 G9	---
		C17 C4		
* 0 *		C1 C2 C3	G2 G3 G6	3
	E1	C23 C24		
		C16 C21 C9		---
	E4	C11 C22		
		C12 C7 C8	G1	
		C18		2
		C19	G7	
	E2			

-1 +				+ (1)
Measr	+Laboratory experiments	+Criteria	+Group	RATING

The science teacher candidates scored five laboratory experiments above the intermediate level, according to the "measr" on the left side of Table 2. Table 3 provides more information on related logit values in greater depth.

Table 3
Logit Values for Three Facets: Laboratory Experiments, Criteria and Group

Laboratory experiments	Logit	Criteria	Logit	Group	Logit
E8	1.43	C13	.93	G5	.33
E7	1.32	C15	.88	G4	.27

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

Laboratory experiments	Logit	Criteria	Logit	Group	Logit
E3	.90	C14	.50	G8	.25
E5	.80	C6	.40	G9	.23
E6	.25	C10	.36	G6	.00
E1	-.07	C5	.21	G3	-.04
E4	-.31	C20	.15	G2	-.05
E2	-.70	C4	.11	G1	-.38
		C17	.06	G7	-.61
		C3	.05		
		C2	-.01		
		C1	-.02		
		C24	-.06		
		C23	-.10		
		C9	-.21		
		C16	-.21		
		C21	-.21		
		C22	-.27		
		C11	-.29		
		C7	-.41		
		C8	-.41		
		C12	-.44		
		C18	-.46		
		C19	-.56		

According to Table 3, there is greater success with the science activity labeled as E8 (logit value: 1.43). Less successful is the science activity with code E2 (logit value: -.70). C19 (logit value: -.56) is the hardest criterion, whereas C13 (logit value:.93) is the easiest. With a logit value of.33, the G5 jury is the most forgiving, and the G7 panel (logit value of -.61) is the harshest.

Laboratory Experiments Performance Analysis

Table 4 displays information regarding the performance analysis of the lab experiments, including the observed average, total score, and logit value.

Table 4
Laboratory Experiments Performance Analysis

Science Activities	Number	Measure	Exp. S.E	**Infit**	ZStd	**Outfit**	ZStd	Total Score	Observed Average
E8	8	1.43	.11	.82	-.7	.83	-.6	1019	4.72
E7	7	1.32	.10	1.02	.1	1.65	2.4	1009	4.67
E3	3	.90	.07	.92	-.5	.65	-2.1	946	4.38
E5	5	.80	.07	.86	-1.1	.88	-.6	925	4.28
E6	6	.25	.05	1.12	1.3	1.11	.9	757	3.50
E1	1	-.07	.05	.97	-.3	.87	-1.3	629	2.91
E4	4	-.31	.05	.89	-1.4	.93	-.5	534	2.47
E2	2	-.70	.06	1.22	1.8	1.13	.7	406	1.88

RMSE (Model) =.07 $\chi^2 = 799.8$ *df* = 7 *p*<.001 Reliability = .99

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Table 4 shows that the dependability coefficient is.99 and the RMSE (Model) is.07. The science activities differ from one another in a quantifiable way. This hypothesis has a separation index of.07, placing it under the fixed effect category. The chi-square test was used to evaluate the reliability coefficient of .99 ($\chi^2 = 799.8$, $df = 7$, $p < .001$). Rejecting the null hypothesis was done. It indicates that the actions differ significantly from one another statistically. The tasks in the qualification sequence are completed in the following order: E8, E7, E3, E5, E6, E1, E4, E2.

Criteria Analysis

The reliability coefficient is .92, and the separation index is 3.46. The standards by which science-related activities are judged diverge significantly. Chi-square was used to test this hypothesis ($\chi^2 = 246.0$, $df = 23$, $p < .001$). Rejecting the null hypothesis was done. These findings indicate that the criteria used to rate science activities differ statistically significantly. Table 5 displays specifics on the criteria measurement analysis, such as the logit value, total score, and observed average.

Table 5
Criteria Measurement Report

Criteria	Meas.	S.E	**Infit**	ZStd	**Outfit**	ZStd	Total score	Obs. Aver.
C13	.93	.15	.96	.0	1.32	.8	327	4.54
C15	.88	.14	1.18	.7	1.27	.7	325	4.51
C14	.50	.12	1.40	1.8	2.22	2.7	302	4.19
C6	.40	.12	1.35	1.7	1.72	1.8	295	4.10
C10	.36	.11	.76	-1.3	.86	-.3	292	4.06
C5	.21	.11	1.28	1.4	1.57	1.6	280	3.89
C20	.15	.11	.95	-.2	.86	-.3	275	3.82
C4	.11	.11	.76	-1.4	.68	-1.1	271	3.76
C17	.06	.11	1.04	.2	.79	-.6	267	3.71
C3	.05	.11	.91	-.4	1.23	.8	266	3.69
C2	-.01	.11	.95	-.2	.95	.0	261	3.63
C1	-.02	.11	.82	-.9	.91	-.2	260	3.61
C24	-.06	.10	1.28	1.5	1.11	.4	256	3.56
C23	-.10	.10	1.04	.2	.79	-.7	253	3.51
C9	-.21	.10	1.08	.5	.94	-.1	242	3.36
C16	-.21	.10	.91	-.4	.74	-.9	242	3.36
C21	-.21	.10	1.09	.5	.79	-.7	242	3.36
C22	-.27	.10	.82	-.9	.66	-1.3	237	3.29
C11	-.29	.10	.81	-1.0	.67	-1.3	235	3.26
C7	-.41	.10	1.00	.0	.89	-.3	224	3.11
C8	-.41	.10	1.05	.3	.95	-.1	224	3.11
C12	-.44	.10	.84	-.8	.85	-.5	221	3.07
C18	-.46	.10	.88	-.6	.65	-1.4	219	3.04
C19	-.56	.10	.91	-.4	.76	-.8	209	2.90

RMSE (Model) = .11 $\chi^2 = 246.0$ $df = 23$ $p < .001$ Reliability = .92

Table 5 shows that the simplest criterion is convenient materials. The criteria are listed from the simplest to the hardest: convenient materials: C13, reliability tested (at least 3 times): C15, step by step movement: C14, well-explained hypothesis: C6, design is for problem solving: C10, hypothesis and problem are parallel: C5, correct measurement calculation: C20, solvable problem: C4, convenient tables/graphics: C17, well-content problem: C3,

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

well-explained problem: C2, qualitative problem: C1, hypothesis based evaluation: C24, well-explained content results: C23, correct controlled variable: C9, correct variables are controlled: C16, correct tables with variables: C21, results explain the problem: C22, changing independent variable: C11, correct dependent variable: C7, correct independent variable: C8, measuring dependent variable: C12, tabled variables: C18, correct units of variables: C19.

Jury Analysis

Table 6 provides information regarding the jury's analysis, including the logit value, total score, and observed average. The jury was composed of science teacher candidates.

Table 6
Group Measurement Report

Group	Nu	Measure	Exp. S.E	**Infit**	ZStd	**Outfit**	ZStd	Total Score	Obsvd Average
G5	5	.33	.07	.95	-.4	.71	-1.5	766	3.99
G4	4	.27	.07	.96	-.3	1.10	.5	755	3.93
G8	8	.25	.07	.77	-2.1	.69	-1.7	749	3.90
G9	9	.23	.07	.95	-.4	.88	-.6	745	3.88
G6	6	.00	.07	1.03	.3	1.58	2.8	694	3.61
G3	3	-.04	.07	.87	-1.1	.86	-.7	685	3.57
G2	2	-.05	.07	1.31	2.5	1.08	.4	683	3.56
G1	1	-.38	.06	1.39	3.1	1.15	.9	602	3.14
G7	7	-.61	.06	.68	-3.2	1.02	.1	546	2.84

RMSE (Model) = .07 χ^2= 191.1 *df* = 8 *p*< .001 Reliability = .95

The result for the reliability coefficient is.95. 4.69 is the group jury separation index. When the hypothesis "there is a difference between severity/leniency of the group jury" was tested using a chi-square test (χ^2=191.1, *df*=8, *p*<.001), the null hypothesis was rejected. Table 6 shows that the group jury with code G5 is the most lenient, while the group jury with code G7 is the severest. Judges are ranked in G7, G1, G2, G3, G6, G9, G8, G4, and G5 in order of leniency to severity.

Jury Bias Analysis

Table 7 presents the group juries, logit values, observed scores, and anticipated scores as part of the bias/interaction report.

Table 7
Bias/Interaction Report

Obs. Score	Exp. Score	Obs. Count	Obs-Exp Average	Group	measr	SciAct.	Proj Measr+
75	105.07	24	-1.25	G7	-.61	E8	1.43
40	82.69	24	-1.78	G2	-.05	E6	.25
64	95.93	24	-1.33	G1	-.38	E3	.90
98	113.08	24	-.63	G6	.00	E7	1.32
71	96.21	24	-1.05	G4	.27	E6	.25
61	84.97	24	-1.00	G5	.33	E1	-.07
97	109.53	24	-.52	G8	.25	E5	.80
38	56.29	24	-.76	G3	-.04	E4	-.31

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

Obs. Score	Exp. Score	Obs. Count	Obs-Exp Average	Group	measr	SciAct.	Proj Measr+
91	103.59	24	.-52	G3	-.04	E5	.80
107	114.03	24	-.29	G6	.00	E8	1.43
40	55.38	24	-.64	G5	.33	E2	-.70
119	110.99	24	.33	G8	.25	E3	.90
119	109.51	24	.40	G1	-.38	E8	1.43
111	98.05	24	.54	G5	.33	E6	.25
109	95.17	24	.58	G8	.25	E6	.25
44	34.41	24	.40	G1	-.38	E2	-.70
101	82.54	24	.77	G4	.27	E1	-.07
104	84.88	24	.80	G6	.00	E6	.25
57	42.15	24	.62	G2	-.05	E2	-.70
70	52.12	24	.75	G1	-.38	E1	-.07
75	55.93	24	.79	G2	-.05	E4	-.31
110	83.09	24	1.12	G3	-.04	E6	.25
120	105.97	24	.58	G3	-.04	E3	.90
120	111.37	24	.36	G4	.27	E3	.90
120	112.02	24	.33	G5	.33	E3	.90
120	110.75	24	.39	G5	.33	E5	.80
120	115.58	24	.18	G5	.33	E7	1.32
120	102.38	24	.73	G7	-.61	E7	1.32
120	113.62	24	.27	G2	-.05	E8	1.43
120	113.70	24	.26	G3	-.04	E8	1.43
120	115.87	24	.17	G4	.27	E8	1.43
120	115.71	24	.18	G8	.25	E8	1.43
120	115.61	24	.18	G9	.23	E8	1.43
86.5	86.47	24.0	.00		Mean (Count:72)		
28.6	26.15	.0	.52		S. D. (Population)		
28.8	26.33	.0	.52		S. D. (Sample)		

Fixed (all = 0) $\chi^2 = 292.8$ *df.*= 72 significance (probability) < .001

Table 7 indicates that some groups could be very strict or very lenient when it comes to science-related activities. For instance, in the science activity, G3 (coded as E4) received 38 points, although the predicted score was 56.29; in the same way, G2 (coded as E4) received 75 points, whereas the expected score was 55,93 points. The expected score for the science activity was 105.07 points, but G7 (classified as E8) only received 75 points. Figure 1 and Table 8 indicate the qualitative results.

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Figure 1
Pre-Service Science Teachers' Views About Peer/Self-Assessment

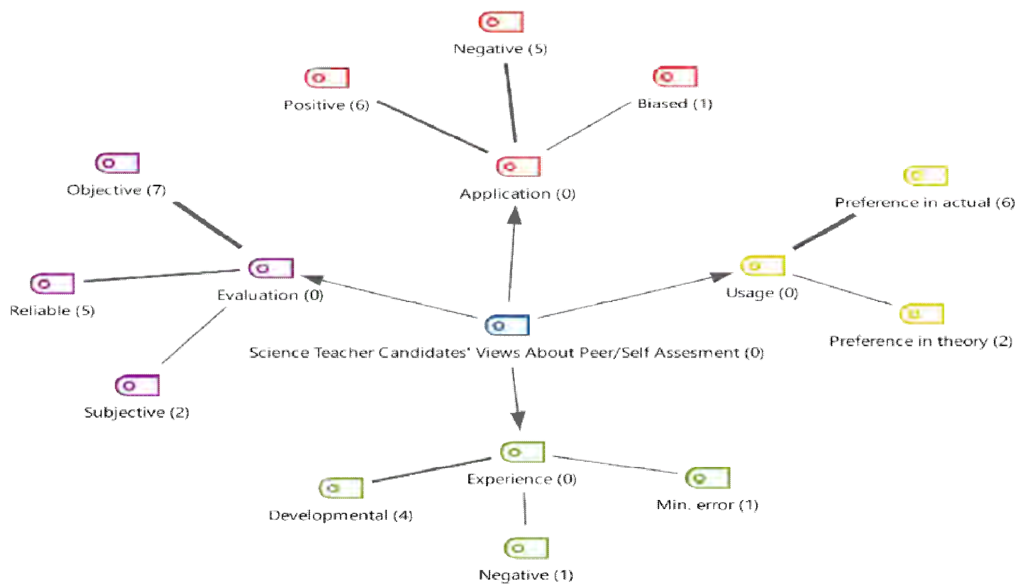

Table 8
Qualitative Results on Science Teacher Candidates' Views about Peer/Self-Assessment

Categories and Codes	Quotes
Evaluation	
• Objective	P6: Yes, we can describe ourselves as a reliable and objective evaluator. When evaluating our own performance objectively, we try to be honest and observe every aspect from a critical perspective. Likewise, we try to evaluate the performance of our colleagues or other teachers objectively. P5: Of course, yes. Since we know what the report our teacher wants is like, we pay attention to it. We evaluate our friends' opinions just as objectively. We never treat anyone biased.
• Reliable	P8: While doing peer self-evaluation, we did not consider our sincerity with anyone and evaluated only based on the experiment report. That's why we think we are objective and reliable. P1: When making an evaluation, I only look at the evaluation sheet, not the person doing it.
• Subjective	P2: We can't say it's very good. Because sometimes we can make comparisons while evaluating, or some criteria may seem inadequate or inadequate to us. We cannot say that it is objective and reliable for sure. P4: We evaluated the report by preparing an experiment in which we converted motion energy into light energy. Some groups were not objective.
Application	
• Positive	P7: On a positive note, we saw our shortcomings through the eyes of our peers because of multiple votes. P3: Our positive aspects are that we maintain our respect for each other. We respect each other and respect your opinions. We distribute the work equally to everyone. There are no under- or over-employments among us. Our negative aspects are almost non-existent when we look at them in general. If we have only one flaw, it would be that we do it slowly.
• Negative	P1: I don't think everyone scores fairly. P6: It can sometimes be difficult to make personal self-criticism and loss of motivation may occur in the process. Additionally, some teachers may feel that the evaluation process is unfair. However, we think that, in general, these practices support the professional growth of educators.
• Biased	P2: The adequacy of the evaluation criteria may vary from person to person. In other words, while the parts of the report are sufficient for me, they may be insufficient for someone else, or vice versa. Sometimes our friends can act biased.

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Categories and Codes	Quotes
Usage	
• Preference in actual	P3: It may vary from subject to subject and grade level, but we plan to use it in general. It is useful because it introduces the student to his strengths and weaknesses. It also teaches us to take responsibility. P7: In my opinion, it should be used because it increases cooperation and interaction in the classroom with peer evaluation. It makes students pay more attention. It helps students realize their strengths and weaknesses.
• Preference in theory	P1: Even if I did peer evaluation, I wouldn't add it to the actual score because I didn't trust the students to keep their emotional side in the background. P2: We can consider using it because students will have experience, but we may obtain a lower evaluation rate because they may think biased.
Experience	
• Developmental	P6: This process helped me improve my teaching skills and guide my students more effectively. P5: Measurement and evaluation, as the literal meaning of the word, is a subject we have mastered. We also had the chance to experience practical lessons. For example, evaluating other groups and us in science teaching or laboratory classes.
• Negative	P7: We didn't have enough time. We couldn't do a detailed review because it was rushed. We could not understand what was expected from us while making the evaluation, so we had difficulty.
• Min. error	P2: We tried to evaluate the items as much as we could, but some items were left between two points, and we gave points to some of them by comparing them with other groups.

Discussion

In the current study, the multi-facet Rasch measurement model was used in the quantitative dimension to examine the opinions of science teacher candidates regarding their ability to prepare experiments based on science process skills, and the MAXQDA-20 software was used in the qualitative dimension to do the same. The simultaneous surfaces employed in the research (experiments based on strictness/generosity of the juries, adequacy of the materials used, and science process skills) were ranked among themselves as a result of the Rasch measurement model. Accordingly, out of eight laboratory experiments based on science process skills, experiment 8 had the highest quality and experiment 2 had the lowest quality. On the other hand, when the items prepared regarding the experiment preparation criteria depend on science process skills were examined, "The materials in the experiment are suitable for the problem and design of the experiment." (C13) and "At least three tests were made to ensure the reliability of the experimental data." (C15) were the easiest items. The most difficult criterion to fulfill was "The units of the dependent and independent variables are written correctly." (C19). When research in the literature was examined, controversial results were encountered in using science process skills in science laboratory experiments. While some studies presented science teacher candidates to be very successful in "Correct units of variables" (Saka, 2019; Koyunlu Ünlü, 2020), some studies presented them as unsuccessful (Govindasamy, Samsudin & Bakar, 2015; Muşlu Kaygisiz, Zirve & Uçar, 2017). Numerous studies have consistently found that students exhibit a deficient proficiency in science process skills (e.g., Aydoğdu, 2015; Irwanto et al., 2017; Özgelen, 2012). Tilakaratne and Ekanayake (2017) found in another study that pupils' basic process skills were poor. Furthermore, Öztürk, Tezel, and Acat (2010) noted that pupils' control over variables and inferential abilities were lacking. Supporting this research finding, Irwanto and Prodjosantoso (2018) posited that science process skills could be developed to determine the dependent and independent variables used in an experiment. When the jury's strictness/generosity knowledge regarding designing laboratory experiments based on science process skills was evaluated, G5-coded jury members displayed the "most generous" behaviour, and G7-coded jury members exhibited the "strictest" behaviour. Supporting this conclusion, various rater characteristics produced statistically significant differences between raters in numerous studies in the literature (Baştürk, 2008; Baştürk, 2010; Köse et al., 2016; Mumpuni et al., 2022; Semerci, 2012; Semerci, 2011; Semerci et al., 2013; Uyanık et al., 2019; Yüzüak et al., 2015).

Within the scope of the qualitative results of the research, participants' positive/negative opinions about application, experiences, reliability/objectivity, and future usage tendencies regarding the peer and self-evaluation process were examined. Teacher candidates saw themselves as reliable and objective evaluators. When evaluating themselves and their peers, they made evaluations according to the evaluation sheet, not

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

bilateral relations. Although they made performance according to predetermined criteria for objective and reliable evaluations, very few juries added their own opinions to the evaluation process and acted subjectively. To support this conclusion, the majority of research studies employing the many-facet Rasch model in the literature (Akın & Baştürk, 2012; Atılgan, 2005; Baştürk, 2008, Baştürk, 2010; Semerci, 2012, Semerci, 2011; Semerci et al. al. 2013; Uyanık et al. 2015) show that raters are not always neutral. Half of the participants stated that there was no negative aspect of the process and that their shortcomings were evaluated and expressed by their friends in the process. However, the other half of the participants stated that they had problems making self-criticism and maintaining their motivation. In addition, the criteria were sometimes inadequate, and the evaluators were biased. Therefore, by going over and editing these items, the assessment forms can be made even better. According to the findings, analysing unexpected reactions might greatly enhance peer and self-evaluation procedures. Most participants stated that using it in their future classes would increase interaction and cooperation among students. It also offers students the chance to improve themselves by showing their weaknesses and strengths. Although participants think that using it theoretically will be effective in increasing students' experience, they believe that using peer evaluation as a score will cause prejudice because it has an emotional dimension. Results from previous studies in the literature were discovered to corroborate the conclusions of this study. Teacher candidates saw themselves as reliable raters when performing peer and self-evaluation, and they also gained experience and had positive opinions about the process (Demir, 2023). Self-assessment and peer assessment have been found to enable participants to collaborate and communicate effectively and support content knowledge development (Fang et al., 2021; Sahin-Taskin, 2018; Tait-McCutcheon & Knewstubb, 2018).

In terms of experience, participants stated that they improved their teaching skills, that they would be more helpful to their students, and that they felt more expert in assessment and evaluation from a developmental perspective. The negative experience was that the process was fast and when there was a difference between two points in the evaluation, they tried to make minimum mistakes by comparing the performances of the groups and giving points. Raters may agree on aspects unrelated to the actual measurement. Therefore, it was essential to gain a deeper understanding of the rating process, including how raters approach the task, the factors they consider, and the characteristics that influence their rating behaviour. Lumley's (2005) model illustrates that raters played a crucial role in the evaluation process. This process was marked by conflict and challenge, as raters brought their unique personalities and past experiences to the rating task. Understanding how these individual traits and histories shape their evaluations was vital to comprehend the impact they have on the outcomes of their ratings. As science teachers tasked with assessing students' science lab experiment performance, they would inevitably judge the students' work, whether it's an experiment, a portfolio, or a written assignment. It is crucial to recognize that the act of judging can introduce undesirable variations or mistakes in the assessment process, which can impact the quality of the students' evaluations (Govindasamy et al., 2015). Therefore, it was important to scrutinize and be aware of the judging behaviour to ensure fair and accurate ratings of students' performances.

Conclusions and Implications

The results of this research may offer encouragement and promise. The study was conducted using both quantitative and qualitative methods to provide compelling and well-founded explanations for the obtained findings. However, a limitation of this research is the small number of teacher assessors. To enhance the validity, studies should be conducted across various schools and locations within the country. Additionally, considering different school categories (such as vernacular, national, and boarding schools) and the location types (urban, semi-urban, and rural) can significantly impact the findings. Another limitation is the limited number of items (24 items) evaluated by the raters focused on a single topic. Future research should expand the number of items and encompass a broader range of topics within the laboratory syllabus for science process skills in science education. Assessments should be conducted over a longer duration to substantiate the findings, ensuring that each examinee has ample opportunity to demonstrate their capabilities. Lastly, the small sample size precludes definitive conclusions about the reliability of rater assessments. Subsequent studies should aim to address this critical issue. Rasch analysis produced a reliability coefficient that resembled both the KR 20 and Cronbach alpha reliability values. When the analysis results and reliability coefficients were evaluated together, the groups scored consistently, and the form prepared within this framework served the purpose.

As a result of all of these, the Multi-Facet Rasch measurement model could be used effectively to evaluate

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

peer groups in science education and objective results could be obtained. At the end of the related research, it can be easily seen that Multi-Facet Rasch Analysis gives the reliable results and provides objectivity throughout the evaluation process. With the help of that, science teacher candidates have been also able to see their level based on the criteria and peer assessments. It can be concluded that in terms of teaching laboratory activities that are based on the science process skills, while some of the science teacher candidates have the necessary qualifications, some of them don't have it. Therefore, it has been observed that the science teacher candidates who have less successful results should be focused more on the following educational processes.

Declaration of Interest

The authors declare no competing interest.

References

Abd-El-Khalick, F., Bell, R. L. & Lederman, N. G. (1998). The nature of science and instructional practice: making the unnatural natural. *Science Education, 82*(1), 417–436.

Abdul Aziz, A., & Masodi, M. S. (2010). Workshop on Rasch analysis: A practical guide to winsteps. *Rasch-Workshop-Booklet*.

Akın, Ö., & Baştürk, R. (2012). The evaluation of the basic skills in violin training by many facets of Rasch model. *Pamukkale University Journal of Education, 31*(1), 175–187. https://dergipark.org.tr/en/pub/pauefd/issue/11112/132860

Aktamış, H., & Ergin, Ö. (2007). Investigating the relationship between science process skills and scientific creativity. *H. U. Journal of Education, 33*(3), 11–23. http://efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/987-published.pdf

Aktamış, H., & Şahin Pekmez, E. (2011). A study of developing scientific process skills inventory towards science and technology course. *Buca Faculty of Education Journal, 30*(30), 192–205. https://dergipark.org.tr/tr/pub/deubefd/issue/25121/265269

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives*. Longman.

Atılgan, H. (2005). Analysis of Special Ability selection examination for music education department using many facets Rasch measurement (İnönü University case). *Eurasian Journal of Educational Measurement, 10*(20), 62–73.

Aydoğdu, B. (2015). The investigation of science process skills of science teachers in terms of some variables. *Educational Research and Reviews Full, 10* (5), 582–594.

Aydoğdu, B. (2017). A study on basic process skills of Turkish primary school students. *Eurasian Journal of Educational Research, 67*(2017), 51–69. https://doi.org/10.14689/ejer.2017.67.4

Aydoğdu, B., Tatar, N., Yıldız, E., & Buldur, S. (2012). The science process skills scale development for elementary school students. *Journal of Theoretical Educational Science, 5*(3), 292–311. https://dergipark.org.tr/tr/pub/akukeg/issue/29346/314037

Azizah, K. N., Ibrahim, M., & Widodo, W. (2018). Process skill assessment instrument: Innovation to measure student's learning results holistically. *Journal of Physics: Conference Series, 947*(1). https://doi.org/10.1088/1742-6596/947/1/012026

Bahşı, A., & Açikgül Firat, E. (2020). The effects of STEM activities on 8th grade students' science process skills, scientific epistemological beliefs and science achievements. *Ondokuz Mayis University Journal of Education Faculty, 39*(1), 1–22. https://doi.org/10.7822/omuefd.616509

Baştürk, R. (2008). Applying the many facet Rasch model to evaluate PowerPoint presentation performance in higher education. *Assessment and Evaluation in Higher Education, 33*(4), 431–444. https://doi.org/10.1080/02602930701562775

Baştürk, R. (2010). Evaluation of scientific research assignments with the multifaceted Rasch measurement model. *Journal of Measurement and Evaluation in Education and Psychology, 1*(1), 51–57. https://dergipark.org.tr/en/pub/epod/issue/5808/77254

Bearman, M., Boud, D., Ajjawi, R. (2020). New directions for assessment in a digital world. In: Bearman, M., Dawson, P., Ajjawi, R., Tai, J., Boud, D. (Eds), *Re-imagining University Assessment in a Digital World*. The Enabling Power of Assessment, vol 7. Springer. https://doi.org/10.1007/978-3-030-41956-1_2

Chang, M. L., & Engelhard Jr, G. (2016). Examining the teachers' sense of efficacy Scale at the item level with Rasch measurement model. *Journal of Psychoeducational Assessment, 34*(2), 177–191.

Çetin, B., & Ilhan, M. (2017). An analysis of rater severity and leniency in open-ended mathematic questions rated through standard rubrics and rubrics based on the SOLO taxonomy. *Education and Science, 42*(18).

Cotabish, A., Dailey, D. & Hughes, G. D., Robinson, A. (2011). The effects of a STEM professional development intervention on elementary teachers' science process skills. *Research in the Schools, 18*(2), 16–25.

Demir, S. (2023). Analysis of peer and self-assessments using the many-facet Rasch measurement model and student opinions. *Journal of Measurement and Evaluation in Education and Psychology*, *14*(3), 266–287.

Duru, M. K., Demir, S., Önen, F. & Benzer, E. (2011). The effects of inquiry-based laboratory applications to preservice science teachers' laboratory environment perceptions, attitudes and scientific process skills. *Marmara University Atatürk Education Faculty Journal of Educational Sciences, 33*(2), 25–44.

Duruk, U., Akgün, A., Dogan, C., & Gülsuyu, F. (2017). Examining the learning outcomes included in the Turkish science curriculum in terms of science process skills: A document analysis with standards-based assessment. *International Journal of Environmental and Science Education, 12*(2), 117–142.

https://doi.org/10.33225/jbse/24.23.641

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Eckes, T. (2005). Examining rater effects in TestDAF writing and speaking performance assessments: A many facet Rasch analysis. *Language Assessments Quarterly, 2*(1), 197–221.

Ergül, R., Şimşekli, Y., Çalış, S., Özdilek, Z., Göçmençelebi, S., & Şanlı, M. (2011). The effects of inquiry-based science teaching on elementary school students' science process skills and science attitudes. *Bulgarian Journal of Science and Education Policy (BJSEP),* 5(1), 48–69.

Fang, J. W., Chang, S. C., Hwang, G. J., & Yang, G. (2021). An online collaborative peer-assessment approach to strengthening pre-service teachers' digital content development competence and higher-order thinking tendency. *Educational Technology Research and Development*, *69*(2), 1155–1181. https://doi.org/10.1007/s11423-021-09990-7

Farsakoğlu, C. F., Şahin, Ç., Karslı, F., Akpınar, A. & Ultay, N. (2008). A study on awareness levels of prospective science. *World Applied Sciences Journal, 4*(2), 174–182.

Fathonah, N., Rahardjo, S., & Prayitno, B. (2018). Analysis of students' science process skill of junior high school in Ngawi district in material classification and its changes. *Advances in Social Science, Education and Humanities Research, 26*(2), 302–305. https://doi.org/10.2991/ictte-18.2018.55

Gill, A. F. (2019). *A study of progression in science process skills during a midwestern public research university's STEM summer camps* [Unpublished doctoral dissertation]. Northern Illinois University, Illinois.

Govindasamy, G. K., Samsudin, M. A., & Bakar, R. A. (2015). To determine the effect of raters on science process skills performance assessment among primary school students. *Sains Humanika*, *7*(1), 43–51. https://doi.org/10.11113/sh.v7n1.597

Hamarat, E., & Arkan, A. (2018). 2023 Eğitim vizyon belgesi'nde gelecek becerileri [Future skills in the 2023 Education vision document]. *Seta Perspektif, 22*(2), 1–7. https://setav.org/assets/uploads/2018/12/222.pdf

Harlen, W. (1999). Purposes and procedures for assessing science process skills. *International Journal of Phytoremediation, 21*(1), 129–144. https://doi.org/10.1080/09695949993044

Hofstein, A. & Lunetta, N.V. (1982). The role of the laboratory in science teaching: neglected aspect of research. *Review of Educational Research, 52(*1) 201–217.

Huppert, J., Lomask, S. M., & Lazarowitz, R. (2002). Computer simulations in the high school: Students' cognitive stages, science process skills and academic achievement in microbiology. *International Journal of Science Education*, *24*(8), 803–821. https://doi.org/10.1080/09500690110049150

Indri, O. W., Sarwanto, & Nurosyid, F. (2020). Analysis of high school students' science process skills. *Journal of Physics: Conference Series, 1567*(3). https://doi.org/10.1088/1742-6596/1567/3/032098

Irwanto, R., & Prodjosantoso, A. K. (2018). Undergraduate students' science process skills in terms of some variables: a perspective from Indonesia. *Journal of Baltic Science Education*, *17*(5), Article 751. https://doi.org/10.33225/jbse/18.17.751

Irwanto, Rohaeti, E., Widjajanti, E., & Suyanta. (2017a). Students' science process skill and analytical thinking ability in chemistry learning. In: *International Conference on Research, Implementation, and Education of Mathematics and Science, Yogyakarta, Indonesia, 1868*(2), 1–5. https://doi.org/10.1063/1.4995100

Ismail, A., Roslan, L., & Adnan, A. N. (2017, November). Assessment on course outcome performance using Rasch measurement model. In *Engineering Education (ICEED), 2017 IEEE 9th International Conference on (pp. 110-113*). IEEE.

Kan, A. (2007). An alternative method in the new educational program from the point of performance-based assessment: Rubric scoring scales. *Educational Sciences: Theory & Practice, 7*(1), 129–152.

Kaya Uyanık, G., Güler, N., Taşdelen Teker, G., & Demir, S. (2017). Investigation of scale development studies conducted in educational sciences published in Turkey by Many-Faceted Rasch Model. *Journal of Measurement and Evaluation in Education and Psychology, 8*(2), 183–199. https://doi.org/10.21031/epod.291367

Köse, İ. A., Sözbir, S. A., & Kalender, C. (2017). Examination of the violin playing skills by means of Rasch model. *Bolu Abant Izzet Baysal University Journal of Faculty of Education, 16*, 2339–2349

Köse, İ. A., Usta, H. G., & Yandı A. (2016). Evaluation of presentation skills by using many facets of Rasch model. *Bolu Abant İzzet Baysal University Journal of Faculty of Education, 16*(4), 1853–1864. https://dergipark.org.tr/en/pub/aibuefd/issue/28550/304600

Koyunlu Ünlü, Z. (2020). Improving pre-service teachers' science process skills and views about scientific inquiry. *Journal of Theoretical Educational Science, 13*(3), 474–489. http://dx.doi.org/10.30831/akukeg.626165

Leat, D., & Nichols, A. (2000). Brains on the Table: Diagnostic and formative assessment through observation. *Assessment in Education: Principles, Policy ve Practice, 7*, 103–121. https://doi.org/10.1080/713613327

Lederman, J. S., & Stefanich, G. P. (2007). Addressing disabilities in the context of inquiry and nature of science instruction. In L. B. Flick ve N. G. Lederman (Eds.) *Scientific inquiry and nature of science*, 55–74. Springer. https://doi.org/10.1007/978-1- 4020-5814-1_4

Linacre, J. M. (1989). *Many-Facet Rasch measurement*. MESA Press.

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch at the Measurement Transactions*, *7*(4), 328.

Linacre, J. M. (1997). Kr-20/Cronbach alpha or Rasch person reliability: Which tells the "truth"? *Rasch Measurement Transactions, 11*(3), 580–1.

MoNE [Ministry of National Education] (2018). Science course curriculum (3rd, 4th, 5th, 6th, 7th and 8th grades). National Education Printing House.

Mumpuni, K. E., Priyayi, D. F., & Widoretno, S. (2022). How do students perform a peer assessment? *International Journal of Instruction, 15*(3), 751–766. https://doi.org/10.29333/iji.2022.15341a

Muslu Kaygısız, G., Benzer, E., & Uçar, M. (2017). Evaluation on preservice science teachers' experimental design related to scientific process skills. *Sakarya University Journal of Education*, *7*(3). 467–483. https://doi.org/10.19126/suje.286360

OECD (2019). OECD Future of Education and Skills 2030 Concept Note. www.oecd.org/education/2030-project

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Özgelen, S. (2012). Students' science process skills within a cognitive domain framework. *Eurasia Journal of Mathematics, Science & Technology Education, 8* (4), 283–292.

Öztürk, N., Tezel, Ö., & Acat, M. B. (2010). Science process skills levels of primary school seventh grade students in science and technology lesson. *Journal of Turkish Science Education, 7*(3), 15–28.

Paine, A. R. (2020). Students' use of science process skills in introductory level biology labs [Unpublished doctoral dissertation]. University of Colorado, Colorado.

Prieto, G., & Nieto, E. (2014). Analysis of rater severity on written expression exam using many faceted Rasch measurement. *Psicologica, 35*, 385–397.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. The University of Chicago Press.

Richardson, M., & Clesham, R. (2021). Rise of the machines? The evolving role of AI technologies in high-stakes assessment. *London Review of Education, 19*(1), 1–14. https://doi.org/10.14324/lre.19.1.09

Rillero, P. (1998). Process skills and content knowledge. *Science Activities, 35*(3), 30–43.

Rowland, P., Stuessy, C. L. & Vick, L. (1987). *Basic science process skills: An in-service workshop kit. [Workshop manual]*. ERIC database.

Sahin-Taskin, C. (2018). Effects of active learning environments supported with self-and peer assessment on pre-service teachers' pedagogical and self-efficacy beliefs. *Asia-Pacific Journal of Teacher Education*, *46*(5), 421–440.

Semerci, Ç. (2011). Analyzing microteaching applications with many-facet Rasch measurement model. *Education and Science*, *36*(161), 14–25.

Semerci, Ç. (2012). The evaluation of students on ideas about the department of computer education and instructional technology (ceit) according to Rasch measurement model. *e-Journal of New World Sciences Academy, 7*(2), 777–784.

Semerci, Ç., Semerci, N., & Duman, B. (2013). Analysis of seminar presentation performances of postgraduate students with many-facet Rasch model. *The Journal of SAU Education Faculty, 25*, 7-22.

Şensoy, Ö., & Yıldırım, H. İ. (2017). Effects of Inquiry Based Learning Approach on Creative Thinking and Scientific Process Skills. *Cumhuriyet International Journal of Education-CIJE, 6*(1), 34–46. https://doi.org/10.30703/cije.321434

Serevina, V., Sunaryo, S., Raihanati, R., Astra, I. M., & Sari, I. J. (2018). Development of Emodule based on problem based learning (PBL) on heat and temperature to improve students' science process skills. *TOJET: The Turkish Online Journal of Educational Technology, 17*(3), 26–36.

Strong, M. G. (2013). *Developing elementary math and science process skills through engineering design instruction* [Unpublished master's thesis]. Hofstra University, New York

Tait-McCutcheon, S., & Knewstubb, B. (2018). Evaluating the alignment of self, peer and lecture assessment in an Aotearoa New Zealand pre-service teacher education course. *Assessment & Evaluation in Higher Education*, *43*(5), 772–785.

Tilakaratne, C. T., & Ekanayake, T. M. (2017). Achievement level of science process skills of junior secondary students: Based on a sample of grade six and seven students from Sri Lanka. *International Journal of Environmental & Science Education, 12* (9), 2089–2108.

Türkmen, H., & Kandemir, E. M. (2011). Öğretmenlerin bilimsel süreç becerileri öğrenme alanı algıları üzerine bir durum çalışması. *Journal of European Education, 1*(1), 15–24. http://eu-journal.org/index.php/JEE/article/view/171/157

Uyanık, G. K., Güler, N., Teker, G. T., & Demir, S. (2019). The analysis of elementary science education course activities through many-facet Rasch model. *Kastamonu Education Journal, 27*(1), 139–150. https://doi.org/10.24106/kefdergi.2417

Uysal, E., ve Cebesoy, Ü. B. (2020). Investigating the effectiveness of design-based stem activities on pre-service science teachers' science process skills, attitudes and knowledge. *SDU International Journal of Educational Studies, 7*(1), 60–81. https://doi.org/10.33710/sduijes.614799

Webb, N. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education, 20*(1), 1–20.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. MESA Press.

Wu, J. (1994). Understanding the nature of science, science process skills, and attitudes toward science of students exposed to different science curricula in Taiwan, Republic of China [Unpublished doctoral dissertation]. The University of Iowa, Iowa.

Yandila, C. D. ve Komane, S. S. (2004). Acquisition of scientific process skills in Botswana general certificate secondary education science. *The Journal of the International Council of Associations for Science Education, 16*(4), 333–344.

Yılmaz, F., & Sözer, M. A. (2018). Examining the 4th grade students' ability to say ―no‖ in the framework of life studies curriculum. *International Journal of Assessment Tools in Education, 5*(3), 443–460.

Yılmaz-Tüzün, Ö., & Özgelen, S. (2012). Preservice science teachers' beliefs about application of science process skills: A case study. *Education and Science, 37*(164).

Yüzüak, A. V., Yüzüak, B., & Kaptan, F. (2015). A many-facet Rasch measurement approach to analyze peer and teacher assessment for authentic assessment task. *Journal of Measurement and Evaluation in Education and Psychology, 6*(1), 1–11. https://doi.org/10.21031/epod.57425

Received: *May 07, 2024* Revised: *June 15, 2024* Accepted: *August 02, 2024*

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Cite as: Hiğde, E., Yüzüak, A. V., Öcal, Z. M., & Aktamış, H. (2024). A many-facet Rasch measurement approach to analyze the prepared science laboratory activities based on science process skills and views of pre-service science teachers. *Journal of Baltic Science Education*, *23*(4), 641–657. https://doi.org/10.33225/jbse/24.23.641

Emrah Hiğde
(Corresponding author)
PhD, Associate Professor, Aydın Adnan Menderes University, Faculty of Education, Science Education Department, Aydın, Türkiye.
E-mail: emrah.higde@adu.edu.tr
Website: https://akbis.adu.edu.tr/personel_detay.asp?sno=A-2297&tur=yayin
ORCID: https://orcid.org/0000-0002-4692-5119

Ahmet Volkan Yüzüak
PhD, Associate Professor, Science Education Department, Faculty of Education, Bartın University, Bartın, Türkiye.
E-mail: volkanyuzuak@bartin.edu.tr
Website: https://fenbilgisi.bartin.edu.tr/akademik-personel/akademik-personel.html
ORCID: https://orcid.org/0000-0002-4712-0259

Zekiye Merve Öcal
Research Assistant, Science Education Department, Faculty of Education Bartın University, Bartın, Türkiye.
E-mail: zocal@bartin.edu.tr
Website: https://fenbilgisi.bartin.edu.tr/akademik-personel/akademik-personel.html
ORCID: https://orcid.org/0009-0002-2722-1199

Hilal Aktamış
PhD, Professor, Aydın Adnan Menderes University, Faculty of Education, Science Education Department, Aydın, Türkiye.
E-mail: haktamis@adu.edu.tr
Website: https://akbis.adu.edu.tr/personel_detay.asp?sno=A-0801
ORCID: https://orcid.org/0000-0003-0717-5770