Research Article https://doi.org/10.12973/eu-jer.12.3.1297

European Journal of Educational Research

Volume 12, Issue 3, 1297 - 1315.

ISSN: 2165-8714
http://www.eu-jer.com/

Measurement of Students' Chemistry Practicum Skills Using Many Facets Rash Model

Melly Elvira[*]
Universitas Negeri Yogyakarta/Universitas Islam Negeri Maulana Malik Ibrahim Malang, INDONESIA

Heri Retnawati
Universitas Negeri Yogyakarta, INDONESIA

Eli Rohaeti
Universitas Negeri Yogyakarta, INDONESIA

Syamsir Sainuddin
Universitas Cokroaminoto Palopo, INDONESIA

Received: December 19, 2022 ▪ Revised: March 4, 2023 ▪ Accepted: March 24, 2023

Abstract: The accuracy of assessing the capabilities of the process and product in chemical practice activities requires appropriate measurement procedures to be followed. It is crucial to identify the components that can introduce bias while measuring student abilities during the measurement process. This study aims to identify the components or criteria used by teachers to assess student performance in practicum activities and analyze the quality of the rubrics developed. The study was conducted with the participation of three raters, 27 high school students, and nine assessment criteria. A quantitative descriptive approach was employed using the many-facet Rasch model (MFRM) analysis for measurement. The results of the MFRM analysis show no significant measurement bias, with data measurement facets fitting the MFRM model. The reliability of all the facets meets the criteria, and the scale predictor functions appropriately. While all students can easily pass four out of nine items, five items can only be partially passed by students. The assessment criteria that require special attention include communication skills, tools and assembly, interpretation, cleanliness, and accuracy when performing practicums. These criteria provide feedback for teachers and students to ensure successful practicum activities. The Discussion section of this study delves into the findings and their implications.

Keywords: *Chemistry practicum, MFRM, performance assessment, process assessment, product assessment.*

To cite this article: Elvira, M., Retnawati, H., Rohaeti, E., & Sainuddin, S. (2023). Measurement of students' chemistry practicum skills using many facets rash model. *European Journal of Educational Research*, *12*(3), 1297-1315. https://doi.org/10.12973/eu-jer.12.3.1297

Introduction

Measurement of students' skills in practicum activities requires an appropriate assessment system so that the teacher can identify the extent of students' understanding and skills (Giammatteo & Obaya, 2018; Ural, 2016). Testing students' skills can be done using a performance assessment instrument, an assessment analysis referring to the results of observations of all student activities during practicum activities (Hensiek et al., 2016). The rubric is an effective measurement tool used to measure student performance, especially in practicum activities (Brookhart, 2013; Wesolowski, 2012). Choosing a rubric to measure student skills is the right choice for the teacher because the rubric can describe the task instructions given, the components to be assessed, and the rater using the rubric (Almarshoud, 2011). The use of rubrics can provide feedback to students on which components are suitable or lacking in practicum activities. This information is valuable to teachers to improve components still lacking in practicum activities carried out by students (Mitchell, 2006). Feedback is provided by the rater based on the predictors of each assessment component. The assessment results use a rubric to accurately assess students' abilities (Wesolowski et al., 2017).

The importance of rubrics in the field of measurement, especially in chemistry practicum activities, is also reviewed by several studies that have developed practicum rubrics, especially in chemistry, to measure student performance in the laboratory (Adams, 2020; Harwood et al., 2020). In Indonesia, teachers do not use any particular rubric to assess students' practicum abilities (Asmorowati et al., 2021). Teachers only use pre-test, post-test, and practicum reports to assess students' skills in practicum activities.

Studying chemistry is not just about learning concepts, but includes the nature of chemistry, scientific practise, scientific inquiry, and the relationship between science, technology, and society (Orgill et al., 2019; Seery, 2020). In practice and scientific inquiry, it includes process and product skills that will be the essential capital for research in the laboratory

[*] **Corresponding author:**
Melly Elvira, Universitas Negeri Yogyakarta/Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia.
✉ mellyelvira.2017@student.uny.ac.id

© 2023 The Author(s). **Open Access** - This article is under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

and the future (Hlukhaniuk et al., 2020). One of the processes and product skills is providing direct experience to students through practical activities in the laboratory (Hlukhaniuk et al., 2020; Tahya et al., 2022). The hope is that students can change their perception of essential things in chemistry. Therefore, process and product skills need to be trained because they play an essential role in chemistry learning (Tahya et al., 2022). The importance of laboratory skills and its role in providing students with direct experience and firsthand experience. The aim of this experience is to change students' perceptions about important things in the field of chemistry (Ghaemi & Potvin, 2021; Pusca et al., 2017).

A rubric was developed in this study, which focuses on nine assessment criteria designed to guide students in preparing themselves and directing their attention to the expected aspects. The hope is that this rubric will also aid students in planning and preparing for practicum assignments by enabling them to gather the necessary tools and materials, as well as to allocate appropriate time and effort. The rubric was created to evaluate chemistry practicum. The construct of this study consists of two assessment dimensions: process assessment and product assessment. Each dimension has specific assessment criteria and codes, which are discussed in the following paragraphs.

The "moving" code represents the criteria for conducting the practicum procedure as a whole. This aspect refers to students' ability to perform each step of the practicum procedure accurately and systematically (Hunter et al., 2003; Maknun, 2015). These steps include preparation before the practicum begins, conducting the practicum according to the predetermined sequence of steps, and concluding the practicum by cleaning the work area and equipment used (Royal Society of Chemistry, n.d.). Through practicum activities, students can gain direct experience in carrying out the practicum procedure and understand how it relates to the theoretical concepts they have learned.

The "Discipline" code includes the criteria for arriving on time before the practicum begins, and is included in the process assessment dimension of chemistry practicum activities. This criterion indicates students' ability to arrive at the laboratory on time before the practicum begins (Asmorowati et al., 2021; Bodner, 2015; Royal Society of Chemistry, n.d.) This assessment is based on aspects of student discipline in managing time and preparing for the practicum. The disciplinary criteria play a crucial role in achieving the objectives of the practicum assignments.

The criteria for utilizing occupational safety and security measures during the practicum align with the "Work Safety" code. This criterion demonstrates the students' ability to implement appropriate safety and security measures during the practicum (Mistry & Gorman, 2020; Royal Society of Chemistry, n.d.). It is crucial to enforce occupational safety and health protocols in the laboratory to prevent any accidents or hazards that may harm the students and the environment.

The criteria for selecting and assembling the practicum equipment align with the experiment to be conducted with the "Assembling" code. This criterion indicates the students' ability to choose and assemble the practicum equipment based on the type of experiment to be conducted (Maknun, 2015; Neamah, 2020; J. S. Wright et al., 2018). The assessment is based on the students' comprehension of the correct utilization of the practicum equipment and the proper execution of the experiment. The appropriate selection and utilization of laboratory equipment are essential in achieving experiment success and avoiding any accidents.

The criteria for demonstrating proper attitude during practicum with the code "Attitude" in chemistry practicum activities refer to the conduct and demeanor of participants while conducting the experiment. The attitude of practicum participants encompasses a positive approach towards the practicum material, active participation, teamwork, and observance of safety protocols and environmental regulations (Hager et al., 1994; Ng, 2019). A good attitude during the practicum is of utmost importance as it influences the final outcome and the success of participants in comprehending the practicum material.

The criteria for precision in implementing the practicum with the code "Accuracy". This criterion pertains to a high level of precision while conducting the experiment. In the context of chemical practicum assessment, precision implies the students' capacity to perform each stage of the experiment with careful attention, accuracy in the selection of chemicals, tools, and practicum procedures (Capozzi et al., 2015; Gürses et al., 2015; Zengele & Alemayehu, 2016). Moreover, accuracy encompasses the precise measurement and recording of data to produce valid and dependable results.

The criterion for cleaning and tidying up the laboratory table is labeled as "Cleanliness". This criterion is associated with the students' ability to clean and tidy up the laboratory table after the completion of the practicum. In the assessment process, "Cleanliness" can be included in a rubric that comprises specific criteria, such as the students' ability to appropriately dispose of chemical waste, clean the practicum tools, and tidy up the laboratory table once the practicum is finished (Neamah, 2020; Royal Society of Chemistry, n.d.). Moreover, conducting a hygiene assessment can motivate students to maintain a clean practicum environment and develop discipline in completing assignments.

The criteria for accurately interpreting experimental data based on the results of observations in laboratory reports is labeled as "Interpretation". This criterion evaluates the students' ability to analyze experimental data correctly and comprehensively and to relate this data to the theory they have learned (Lunardi et al., 2021). The assessment of the interpretation of experimental data is highly significant as it can produce meaningful information and support research results. Accurately interpreting experimental data can lead to strong conclusions, assist in identifying errors in the experimental process, and enhance research credibility (Asmorowati et al., 2021; Lunardi et al., 2021). Hence, the students' ability to interpret experimental data precisely is a crucial indicator in evaluating the quality of research results.

The criterion for assessing the ability of practicum participants to present experimental observations in a clear and systematic way in the form of a written report is labeled as "Communication". In preparing the report, practicum participants are expected to organize the results of observations into a logical and coherent sequence, and provide brief and precise explanations for the results of observations. Communication skills in report writing are essential in expressing ideas and thoughts in writing, in a language that is easily comprehensible to readers (Asmorowati et al., 2021; Hensiek et al., 2016). Additionally, a well-written report must include complete and accurate information and be written using proper grammar.

In general, Process Assessment evaluates the quality of practicum implementation, while Product Assessment evaluates the quality of practicum result reports. Using rubrics based on this construct is expected to help teachers evaluate student performance objectively and provide appropriate feedback, enabling students to improve their skills in the future (Reigosa & Jiménez-Aleixandre, 2007).

Rubric development requires more time and focus (Mitchell, 2006). The rubric development stage begins with setting the assessment objectives and determining the components of the assessment targets considered necessary in practical activities (Janssen et al., 2015). One type of rubric used in measuring skills is an analytical rubric (Chukwuere, 2021). Analytical rubrics require predictors to be the benchmark for assessment. This predictor usually describes the abilities/skills expected by students in certain activities (Beyreli & Ari, 2009; Chukwuere, 2021).

Instrument reliability is essential in compiling rubrics (Harmey et al., 2019; Johnson et al., 2019; Yamanishi et al., 2019). A good rubric at least goes through the stages of expert review and empirical testing (Basturk, 2008; Galti et al., 2018). Expert review is a consideration for items essential to measuring students' practicum abilities in the laboratory (Sainuddin et al., 2022). Meanwhile, empirical trials were used to determine to what extent the rubric developed appropriately measures students' abilities in practical activities (Bennett et al., 2016; Giammatteo & Obaya, 2018; Seery et al., 2017). Empirically, we can determine the reliability of all facets involved using rubrics such as assessment items, students, and raters.

Based on a previous discussion on the use of measuring instruments in assessing students' practicum abilities in the laboratory, the authors attempted to measure students' practicum skills by developing a practicum assessment rubric. This rubric focuses on the evaluation of processes and products. The criteria used in this rubric include the ability of students to perform practical procedures, discipline, carry out work safety, use tools/materials, attitude, precision, cleanliness, interpretation, and communication in the practicum. Through the rubric criteria developed, the data obtained can be used as a reference by the teacher to assess student performance in practical activities in the laboratory. The quality of the developed rubric was analyzed using the Many Facet Rasch model (MFRM) approaches. This analysis can show the quality of the rubric produced through the data obtained from the students who do the practicum, the number of items and the number of raters involved in the assessment. The quality of the resulting instrument does not necessarily result in a conclusion. However, other aspects can also be studied more deeply, such as the number of raters and students. An unbalanced assessment between raters may cause an instrument that is not considered good. Therefore, through the MRFM analysis, the authors can examine more deeply the quality of the instruments used in measuring the skills of the chemistry practicum. Thus, the study aims to identify what criteria / components teachers need to consider to assess student performance in practical activities in the laboratory and to analyse the quality of the rubrics that have been developed using the MFRM approach.

Methodology

Research Design

The study uses a quantitative descriptive approach to develop a rubric for assessing students' chemistry practicum activities in the laboratory. This study aims to determine the quality of the assessment criteria used to measure students' skills in a practicum in the laboratory.

Research Participant

The participants in this study were 27 high school students in class XII majoring in science consisting of 10 males and 17 females aged 17-18 years. The students who were assessed came from several different class groups. The sampling was done by a random sampling technique. Each class randomly selected nine people to represent their class in practicum activities.

Instrument

The instrument used in this study is an assessment rubric developed by the researcher based on relevant literature, including the "moving" criteria (Hunter et al., 2003; Maknun, 2015); the "Discipline" criteria (Asmorowati et al., 2021; Bodner, 2015; Royal Society of Chemistry, n.d.); the "Work Safety" criteria (Mistry & Gorman, 2020; Royal Society of Chemistry, n.d.); The "Assembling" criteria (Maknun, 2015; Neamah, 2020; J. S. Wright et al., 2018). the "Attitude" criteria (Hager et al., 1994; Ng, 2019); the "Accuracy" criteria (Capozzi et al., 2015; Gürses et al., 2015; Zengele & Alemayehu, 2016). the "Cleanliness" criterion (Neamah, 2020; Royal Society of Chemistry, n.d.); the "Interpretation" criteria (Lunardi et al., 2021); and the "Communication" criterion (Asmorowati et al., 2021; Hensiek et al., 2016). The rubric consists of nine items that are scored using a maximum score of 4 and a minimum score of 1 for each criterion. This assessment rubric will assist teachers and students in measuring and improving the practical skills of students in performing accurate and safe acid-base titration. Table 1 shows examples of item criteria and a scoring rubric for assessing acid-base titration laboratory work.

Table 1. Items and Scoring Rubric for The Assessment

Criteria	Score 4	Score 3	Score 2	Score 1
Arrival time before the practicum test activity begins.	□ Arrives 10 minute before the start of the practical	□ Arrives on time or with a delay of less than 10 minutes	□ Arrives with a delay of more than 10 minutes	□ Absent during the practical
Attitude during the practicum.	□ Demonstrates a highly positive and enthusiastic attitude	□ Demonstrates a positive attitude but lacks enthusiasm	□ Demonstrates a negative attitude and lacks enthusiasm	□ Demonstrates a negative and unenthusiastic attitude
Writing a report in accordance with the observations made during the practicum.	□ Writes a very clear and systematic report	□ Writes a report that is clear and systematic enough	□ Writes a report that is unclear and not systematic enough	□ Does not write a report or writes an incorrect report

Before the instrument was used, a draft of the instrument was given to 13 experts to assess the validity of each item in representing the construct. The 13 experts involved in this study consisted of 3 psychometricians, 5 chemistry professors, and 5 chemistry teachers with at least 10 years of teaching experience. Afterwards, the validated rubric was empirically examined. Experts have provided feedback on the developed instrument, which can be summarized as follows: the predictor section is concise and clear for evaluators to understand. The writing has also been corrected based on proper Indonesian grammar and spelling. Additionally, it is important to ensure proper waste management practices to allow students to dispose of waste correctly. This instrument is expected to provide information and feedback on students' abilities, improve students' skills in conducting chemistry experiments, and serve as an evaluation tool for chemistry teachers.

Data Collection

The objective of the chemical laboratory activity in this study is to ensure that students comprehend the theoretical concepts and are competent in applying them in practical experiments. In this study, the evaluation activity was carried out through an acid-base titration laboratory to assess the students' proficiency in executing the practical experiment. The acid-base titration concept pertains to the determination of the concentration of an acid or a base solution by utilizing a standard solution with a known concentration. Acid-base titration is performed by the addition of a standard solution of a base to an acid solution, or vice versa, until the endpoint or the color change of the indicator is reached. This concept is a critical aspect of chemical analysis, as it enables precise and accurate measurement of the concentration of acid or base solutions. Moreover, the acid-base titration laboratory also develops students' skills in handling laboratory apparatus safely and properly and in generating precise and methodical laboratory reports.

Before collecting data through observation activities, the researcher explained in advance the use of the instrument to three teachers who acted as raters two weeks before the practicum activities. A week before the actual practicum activities is carried out. The teacher can use the rubric for other practicum activities to find out which part of the rubric needs to be corrected.

This study involved three chemistry teachers as raters, two male and one female. Chemistry teachers' qualifications are chemistry teachers who have taught chemistry for at least ten years. Students and raters were disguised in the study or analysis to maintain confidentiality. Three raters jointly observed 27 students who completed chemistry practicums in the laboratory.

The teacher explained the practicum activities for 15 minutes. Afterwards, the students carried out the activities according to the procedures outlined in the practicum module. While the activities were ongoing, the teacher independently assessed the students' ability to carry out the practicum. Once the activities were completed, the students were asked to submit a report, which was collected two days later. The reports were then evaluated by a rater using a rubric to assess the quality of the practicum product.

Data Analysis

Many-Facet Rasch Model (MFRM) is an analysis method used to measure the quality and reliability of a measurement instrument (Eckes, 2015; Linacre, 1994b, 2018). The steps typically involved in MFRM analysis (Eckes, 2015) include: 1) preparing the measurement instrument and determining the characteristics to be measured, 2) selecting the raters who will do the measurement, 3) determining the sample to be measured, 4) conducting MFRM analysis, and 5) evaluating the analysis results and determining necessary improvements. MFRM analysis is used to ensure that the measurement instrument used in research is of good quality and can be relied upon to measure the desired characteristics. This is important in research to ensure that the resulting data can be properly interpreted and used as a basis for making accurate decisions.

By using MFRM, factors that influence the assessment results, such as rater, student, and item characteristics, can be identified (Eckes, 2015). This can help in determining necessary improvements in the measurement instrument, such as providing training to improve rater reliability or revising unclear or irrelevant criteria items (Eckes, 2015; Linacre, 1994a). In the context of using a 9-item rubric involving 3 raters and 27 12th grade students in an acid-base titration practical experiment, MFRM analysis is used to ensure the quality of the assessment instrument. Thus, MFRM analysis can ensure that the assessment conducted on the acid-base titration practical experiment has good quality and can be relied upon to provide accurate and precise information about the students' ability in conducting the experiment.

To perform MFRM analysis, several steps are required. First, the measurement instrument must be prepared, and the characteristics to be measured must be determined. Then, raters who will conduct the assessment and samples that will be measured must be selected. The next step is to perform the MFRM analysis using Facets software. In this study, three facets were used: students, raters, and items. The severity (C) of the rater was tested using MFRM in this study. This severity is considered as the estimated probability that students (n) will give answers to item (i), the threshold for the category (k) for rater (j), which is expressed in equation 1. The results of the MFRM analysis were tabulated in MS. Excel, and the findings were evaluated to determine necessary improvements to the measurement instrument. Overall, MFRM analysis is a crucial method for ensuring the quality and reliability of a measurement instrument and can provide accurate and precise information in various research contexts.

$$\rho_{nikj} = \frac{e^{\left(\beta_n - \delta_i - F_k - C_j\right)}}{1 + e^{\left(\beta_n - \delta_i - F_k - C_j\right)}} \qquad (1)$$

With:

$\beta_{n\_}$	= Function of the student's ability
δi	= Difficulty level of rubric items
Fk	= threshold level of difficulty
C_j	= severity of the rater

Through MFRM analysis, it is hoped that there will be no bias in the assessment. He et al. (2013) state that it is possible to find sources of bias by analyzing the interaction between raters, items, and students. Equation 1 shows that students' practicum activities are assessed based on specific criteria by the rater, a multi-rater function consisting of the ability or skills/skills of students carrying out practicum activities, rater severity, and difficulty of item criteria. All aspects are described based on independent parameters with an interval scale (logit) (Linacre, 2018). The parameter's precision level is described by each facet's standard error of measurement (SEM). In addition, Fair average measures are presented, which show the average score independent without being influenced by the rater's facet and the level of difficulty of the item scale (Linacre, 1994b, 2018).

Separation reliability and separation index

Reliability in the MFRM analysis was estimated for all facets. Reliability is the ratio of the actual score to the observed score for all aspects or facets, such as raters, items, and students. The reliability index ranges from 0 to 1. High reliability indicates a wide distribution of each aspect (Morgan et al., 2014; Weigle, 1998; Yan, 2014). The interpretation of separation reliability varies depending on the facets considered (Eckes, 2015). For examinees, these statistics provide information on how effectively the criteria differentiate between test takers based on their ability level. In other words, examinee separation reliability indicates the extent to which ability measures differ among test takers. Thus, achieving high reliability of examinee separation is a desirable goal (Eckes, 2015; Linacre, 2018).

In contrast, the interpretation of separation reliability is significantly different for raters (Eckes, 2015; Linacre, 2018). If raters within a group have similar severity levels, then the reliability of rater separation will approach 0. From the standard approach to rater variation, achieving low reliability of rater separation is desirable as it demonstrates that raters can be interchanged. However, if raters within a group have vastly different levels of severity, then the reliability of the split raters will approach 1, indicating how different the tightness measurements are. Therefore, it is essential to clearly distinguish between these two types of reliability indices. Therefore, some aspects, such as students and items, are expected to have a high-reliability index, while rater aspects are expected to have low reliability.

Reliability can also be expressed as a separation that shows the number of students' ability levels that are statistically different in the data distribution (Fisher, 2007; Wright, 1996). The dividing index ranges from zero to infinity. The minimum separation index for students' ability is 2.5 (Deviana et al., 2020; Fisher, 2007). Output reports a chi-square test for each facet which shows statistical differences in the level of student ability, rater severity, and item scale difficulty. A significant probability value ($p < .05$) indicates that the facets have the same level for each characteristic (Aryadoust, 2016),

Statistical fit

The output examines the measurement deviation of each facet using the infit statistic mean square (IMS) and outfit mean square (OMS) (Aryadoust, 2016; Linacre, 2002). IMS is a weighting index that provides sensitive information on student performance on items (assessment criteria) at or close to a particular student's ability level. OMS provides accurate information regarding outliers or student performance on items far from the test taker (i.e., erratic patterns on items that are too easy or too difficult for students). The expected IMS or OMS value is 1. So, if the IMS or OMS value is 1.2, it indicates that there is more variation than predicted by Rasch, up to 20% of the expected score (Aryadoust, 2016). There is no fixed standard for IMS values; some experts use a range of 0.6 – 1.4 (Bond et al., 2020), and some use a more stringent standard, namely, in the range of 0.8 – 1.2 (Linacre, 1994b).

Rating scale functionality

The functioning of the rating scale can be observed from the calibration of the Rasch-Andrich threshold (Linacre, 2018). The average size of the categories must increase monotonically from low to higher levels, with the distance between adjacent categories ranging from 1.5 – 5 logit. The OMS index is expected to be 0.5 to 1.5 (Linacre, 2002).

Results

Proof of Content Validity

The initial study to develop the rubric for assessing chemical laboratory activities was measured using process and product assessment dimensions. As explained in the introduction, the instrument, which consists of 9 initial assessment items, is then reviewed for readability by a measurement expert. Experts provide input about the content of each item and see how far the item can represent the construct (Sainuddin et al., 2022). The experts included two other chemistry lecturers, nine chemistry teachers, and two educational measurement experts. Aiken's V (Aiken, 1985) was used to calculate validity using equation 2.

$$V = \frac{\sum s}{\left[n\left(c-1\right)\right]} \qquad (2)$$

where:

S = r – lo

n = number of rater/ experts

lo = the lowest score of validity (e.g., 1)

c = the highest validity rating score (e.g., 5)

r = number given by the rater

Based on the critical value criteria for the value of the validity coefficient, the critical value of the valid item for 13 raters and five category ratings (c = 4) at a significant level of .01 is .77. Table 2 shows that all items have a validity index more significant than the critical value (.77), so it can be concluded that all items and rubric predictors developed to measure student practicum performance are content-valid. The good rubrics are then revised on input from experts. The final instrument is ready for use at the empirical trial stage.

Table 2. Dimensions, Items, and Aiken's Validity Coefficient

Rating Dimension	Assessment criteria	Code	V (Min .77)
Process Assessment	Carry out all practical procedures	1. Moving	0.821
	Arrival time before practicum test activity begins	2. Discipline	0.846
	Using security and safety attributes during practicum according to procedures	3. Work safety	0.769
	Select and assemble practicum equipment according to the experiments to be carried out	4. Assembling	0.821
	Attitude during practice	5. Attitude	0.821
	Accuracy in practical implementation	6. Accuracy	0.795
	Clean and tidy up the lab table	7. Cleanliness	0.821
Product Rating	Interpret experimental data accurately based on the results of observations in the practicum report	8. Interpretation	0.821
	Write a report according to the observations made during the practicum	9. Communication	0.821

Note. V = Aiken Validity Index

Descriptive Statistics

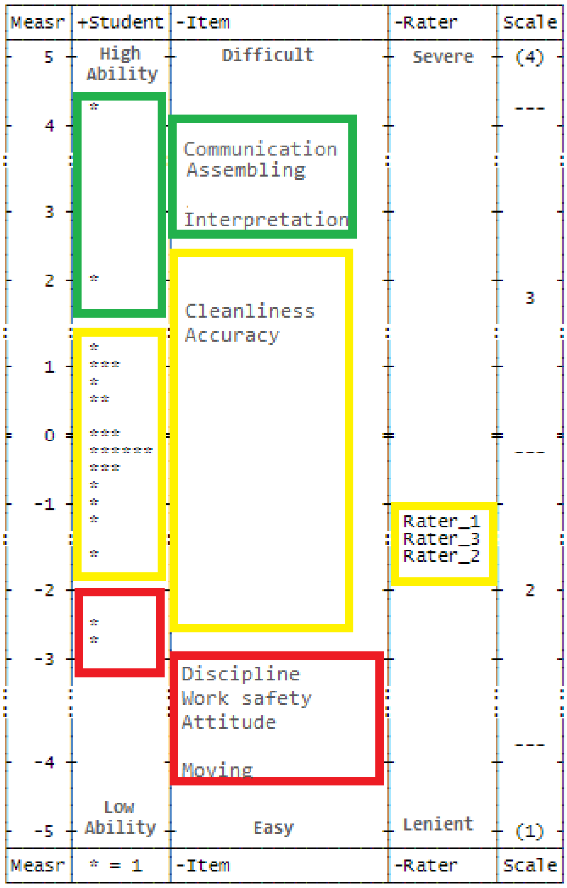
Table 3. Item Mean, Standard Deviation, Skewness, and Kurtosis

Item	Average	SD	Skewness	Kurtosis
1	3.70	0.46	-0.91	-1.20
2	3.56	0.50	-0.23	-2.00
3	3.53	0.50	-0.13	-2.03
4	1.85	0.69	0.67	0.97
5	3.53	0.50	-0.13	-2.03
6	2.51	0.57	0.59	-0.63
7	2.52	0.63	-0.07	-0.20
8	2.11	0.63	-0.09	-0.47
9	1.86	0.67	0.16	-0.72

Note. SD = standard deviation

Table 3 shows the mean, standard deviation, Skewness, and Kurtosis items for assessing student practicum activities in the laboratory. Item number 4 is the most difficult item, with an average rating of 1.85 and a standard deviation of 0.69. On the contrary, the most accessible item is item number 1, with an average of 3.70 and a standard deviation of .46. The distribution of the data shows that the data tends to be expected. This condition can be seen from the value of the skewness coefficient, which is in the score range from -0.91 to 0.67, and the kurtosis coefficient, which has a range from -2.03 to 0.97.

Many - Facet Rasch Model (MFRM) Analysis



```
|Measr|+Student|-Item          |-Rater  |Scale|
|-----+--------+---------------+--------+-----|
|   5 +  High  +   Difficult   + Severe + (4) +
|     | Ability|               |        |     |
|     | *      |               |        | --- |
|   4 +        +               +        +     +
|     |        |Communication  |        |     |
|     |        |Assembling     |        |     |
|   3 +        +               +        +     +
|     |        |Interpretation |        |     |
|   2 + *      +               +        +     +
|     |        |Cleanliness    |        |  3  |
|     |        |Accuracy       |        |     |
|     | *      |               |        |     |
|   1 + ***    +               +        +     +
|     | *      |               |        |     |
|     | **     |               |        |     |
|   0 * ***    *               *        *     *
|     | ****** |               |        | --- |
|     | ***    |               |        |     |
|     | *      |               |        |     |
|  -1 + *      +               +        +     +
|     | *      |               |Rater_1 |     |
|     |        |               |Rater_3 |     |
|     | *      |               |Rater_2 |     |
|  -2 +        +               +        +  2  +
|     | *      |               |        |     |
|     | *      |               |        |     |
|  -3 +        +               +        +     +
|     |        |Discipline     |        |     |
|     |        |Work safety    |        |     |
|     |        |Attitude       |        |     |
|     |        |               |        | --- |
|  -4 +        +Moving         +        +     +
|     |  Low   |               |        |     |
|  -5 + Ability+     Easy      + Lenient+ (1) +
|-----+--------+---------------+--------+-----|
|Measr| * = 1  |-Item          |-Rater  |Scale|
```

Figure 1. Variable Map of Facets.

Variable Map

Figure 1 shows the calibration of the three facets shown on the variable map or the Wright map. The first column describes a logit that ranges from interval (-5, 5). Each student is symbolized by a star (*). The second column shows students' abilities based on their skills/skills in conducting practicum activities; each student is assessed by three people in the fourth column based on the nine assessment items shown in the third column. Students' abilities can be categorised into three ability groups, namely: the high group (2 students) has a score range of more than 2.0 logits, the medium ability group (23 students) is in the range of -2.0 to 2.0 logits, and the low ability group (2 students) is on a score of less than -2.0 on the logit.

The difficulty of the criteria or items in the third column shows three items with a high category of more than 2.0 logits, two items with a moderate category in the range of -2.0 to 2.0 logit, and four easy category items less than -2.0 logit. Column four shows the severity level of the rater, with the most severe at the top and the most lenient at the bottom of the logit. The results show that the rater has a close distribution of raters, showing a homogeneous rating tendency. A severity distribution ranging from -1.39 to -1.13 logit shows that the raters do not differ much in severity. There are no significant differences in severity between the raters.

Reliability and Data Fit

Table 4 displays the psychometric estimates of the rater component. The rater separation index was .00, and the reliability separation was .00. These scores indicate that the distribution of rater severity is relatively homogeneous, suggesting that there are no significant differences in severity between raters. This finding is also supported by an agreement metric, namely Exact agreements = 86.6%, which exceeds the expected agreements of 56.8%. Moreover, the probability sci-square value of .38 is greater than .05, indicating that the rater severity does not differ significantly.

Table 4. Rater Measurement Report

Raters	Total Score	Total Count	Observed Average	Fair (M) Average	Measures	SE	Infit MnSq	Infit ZStd	Outfits MnSq	Outfit ZStd
Raters_3	672	243	2.77	2.87	-1.13	0.13	0.90	-1.20	0.88	-0.90
Rater_1	680	243	2.80	2.89	-1.27	0.13	0.94	-.70	0.93	-0.40
Rater_2	687	243	2.83	2.91	-1.39	0.13	1.10	1.10	1.07	.50
Reliability = .00; Separation =0 .00; exact agreements = 86.6% Expected agreements = 56.8% (χ 2 = 1; df = 1; p = .38)										

Note. MnSq = mean square. SE = Standard error of measurements. ZStd = z standardized.

The severity of the rater assessing students is shown in the measure column, which shows that rater 3 is relatively severe, while rater 2 is relatively more lenient. In general, all raters have a severity level that tends to be moderate, with a score below -1 on the logit. The IMS severity rater is close to 1 (0.90 to 1.10) logit, which means that the rater is fit and meets the expectations of the MFRM model. Meanwhile, OMS also has data that tend to be stable in the 0.88 – 1.07 logit, indicating no rater value outside the model's expectations.

Table 5 shows the statistical estimate of the student component. The student separation index is 3.24, and the reliability is .91. This index shows that the student facet has three levels with sufficient spread, which shows that students' abilities tend to be diverse or heterogeneous. The ability of students to carry out practical activities can be divided into three levels: high, medium, and low. This condition can also be seen on the variable map (Figure 1).

The student's ability is shown in the measure column, which shows the student's ability to be in the range of 4.26 (highest ability = student 12) to -2.86 (lowest ability = student 8) logit. The STI index ranged from an interval of 1.79 (student 12) to 0.26 (student 13). Based on the acceptance criteria for IMS and OMS set at the interval of 0.5 to 1.5, there are five students whose scores do not fit the model: student 12, student 18, student 6, student 4, and student 13. Meanwhile, based on the OMS index, there are four students whose scores are outside the acceptance interval: student 14, student 20, student 4, and student 13.

Table 6 shows the estimates of the statistical components of the assessment criteria or items. The item separation index is 13.66, and the reliability is .99. This index shows that the item facet has 13 levels with a very widespread, this shows the difficulty of the items is very diverse or heterogeneous, the difficulty of the items used to assess practicum activities can be divided into 13 levels, but visually on the variable map, it can be grouped into three levels, namely difficult, medium, and easy groups, see Figure 1.

Table 5. Student Measurement Report

Student	Total Score	Total Count	Observed Average	Fair (M) Average	Measures	SE	infit MnSq	infit ZStd	Outfit MnSq	Outfit ZStd
Student 12	96	27	3.56	3.80	4.26	0.55	1.79	1.80	1.12	0.50
Student 14	88	27	3.26	3.29	2.08	0.48	0.83	-0.40	0.48	-1.00
Student 18	84	27	3.11	3.11	1.26	0.43	1.53	1.50	1.45	1.20
Student 6	83	27	3.07	3.08	1.07	0.42	1.64	1.80	2.13	2.70
Student 19	83	27	3.07	3.08	1.07	0.42	0.96	0.00	1.46	1.30
Student 20	83	27	3.07	3.08	1.07	0.42	0.58	-1.50	0.47	-1.90
Student 1	81	27	3.00	3.02	0.73	0.40	0.98	0.00	0.82	-0.50
Student 27	80	27	2.96	2.99	0.57	0.40	0.61	-1.60	0.61	-1.60
Student 7	79	27	2.93	2.97	0.42	0.39	1.10	0.40	1.00	0.00
Student 2	77	27	2.85	2.91	0.12	0.38	1.45	1.70	1.38	1.50
Student 15	77	27	2.85	2.91	0.12	0.38	1.28	1.10	1.23	1.00
Student 9	76	27	2.81	2.89	-0.02	0.38	0.93	-0.20	1.01	0.10
Student 16	75	27	2.78	2.86	-0.16	0.37	0.98	0.00	0.96	0.00
Student 21	75	27	2.78	2.86	-0.16	0.37	0.97	0.00	0.95	-0.10
Student 23	75	27	2.78	2.86	-0.16	0.37	0.93	-0.20	0.92	-0.30
Student 3	74	27	2.74	2.83	-0.30	0.37	1.02	0.10	1.01	0.00
Student 11	74	27	2.74	2.83	-0.30	0.37	0.95	-0.10	0.94	-0.20
Student 22	74	27	2.74	2.83	-0.30	0.37	0.88	-0.50	0.87	-0.60
Student 17	73	27	2.70	2.80	-0.43	0.37	1.37	1.60	1.35	1.60
Student 26	73	27	2.70	2.80	-0.43	0.37	1.20	0.90	1.17	0.80
Student 25	72	27	2.67	2.77	-0.57	0.37	0.83	-0.70	0.84	-0.70
Student 10	70	27	2.59	2.71	-0.84	0.37	1.13	0.60	1.12	0.60
Student 4	69	27	2.56	2.68	-0.97	0.37	0.49	-2.80	0.49	-2.90
Student 5	67	27	2.48	2.60	-1.25	0.38	0.73	-1.20	0.76	-1.00
Student 24	65	27	2.41	2.52	-1.54	0.38	0.64	-1.50	0.69	-1.20
Student 13	59	27	2.19	2.23	-2.50	0.42	0.26	-3.20	0.23	-3.30
Student 8	57	27	2.11	2.13	-2.86	0.43	0.64	-1.10	0.51	-1.50

Reliability = .91; Separation = 3.24; (χ 2 = 237; df = 26; p < .01)

Note. MnSq = mean square. SE = Standard error of measurements. ZStd = z standardized.

The difficulty level of the items is quite diverse, displayed in the measure column showing the difficulty of the items in the range of 3.87 (difficult = item 4) to -4.04 (easy = item1) logit. The items of the STI index were in the interval 1.53 (item 8) to .49 (item 6). Based on the acceptance criteria for IMS and OMS set at intervals of 0.5 to 1.5, there are two items whose values do not fit the model: item 8 and item 6.

Nine item criteria were developed and valid based on expert judgment. Unfortunately, some criteria still need to be considered in practical activities. These components are the ability of students to assemble practicum tools, how students interpret data correctly from the results of the practicum, how to communicate the results of the practicum in reports, and answering teacher questions during the practicum of practicum activities. Meanwhile, the students' accuracy component (item 6) in measuring and observing during the practicum is also the cleanliness (item 7) of the students shown in the practicum activity, which also requires attention. This information is based on the finding that only two students can carry out these criteria while the rest 25 students failed this component (see Figure 1). Accuracy is essential in measurement and observation activities. Careful students will produce valid data records so that the results reported are by experimental conditions and will affect the conclusion. The cleanliness of the table (item 7) practicum at first glance is usually considered trivial. However, the results of the assessment show that the majority of students need to meet this criterion. One of the essential criteria in practicum activities is the cleanliness and tidiness of students during practicum. Clean students will prepare practical tools and materials neatly and safely dispose of the rest of the practicum.

In general, students can apply discipline (item 2) during the practicum, comply with work safety (item 3), carry out a series of practicums (item 1) according to the instructions in the module and have a good attitude (item 5) when carrying out the practicums. These four criteria need to be achievements for students in practicum activities. Even though they are included in the easy category, they should not be underestimated in practicum activities so that practicum activities carried out by students can run well and as expected.

Table 6. Measurement Items Report

Items	Total Score	Total Count	Observed Average	Fair (M) Average	Measures	SE	Infit MnSq	Infit ZStd	Outfits MnSq	Outfits ZStd
4	150	81	1.85	1.85	3.87	0.21	1.00	0.00	1.02	0.10
9	151	81	1.86	1.87	3.82	0.21	1.43	2.50	1.39	2.20
8	171	81	2.11	2.11	2.92	0.21	1.53	2.90	1.53	2.90
6	203	81	2.51	2.53	1.50	0.21	0.49	-4.60	0.47	-4.60
7	204	81	2.52	2.54	1.46	0.21	0.93	-0.50	0.97	-0.10
3	286	81	3.53	3.56	-3.14	0.24	0.99	0.00	0.95	-0.10
5	286	81	3.53	3.56	-3.14	0.24	0.73	-2.50	0.64	-1.50
2	288	81	3.56	3.59	-3.26	0.24	0.64	-3.40	0.56	-1.90
1	300	81	3.70	3.76	-4.04	0.27	0.90	-0.60	1.13	0.40

Reliability = .99; Separation = 13.66; (χ 2 = 1643; df = 8; p < .01)

Note. MnSq = mean square. SE = Standard error of measurements. ZStd = z standardized

Accurate valuation estimates are crucial for producing valid and reliable measurements. The MFRM analysis can provide a detailed understanding, down to the individual level, of biases or unexpected responses from assessments, such as overestimation or underestimation. Table 7 illustrates the rater's unexpected responses to students using specific items. Each row provides information on how the rater's score for students using certain criteria deviates from the expected score of the MFRM model. Table 7 reveals that out of a total of 759 ratings, 41 ratings show rating bias. The first row in Table 7 shows that rater 2 undervalued student 19 on item 1 (moving) by giving a score of 3, while the expected value of the model is 3.9 or approximately 4. Similarly, row 9 shows that rater 1 overvalued student 19 on item 7 (cleanliness), giving a score of 4, while the expected value of the model is 2.8 or around 3. This explanation applies to rows 1 through 41.

The list of unexpected responses from Table 7 shows the consistency of each rater assessing certain students using specific criteria or items. Based on the summary of the top 3 unexpected responses, it is known that overall, 3.02% of rater ratings are undervalued, and 2.61% of rater's assessment is overvalued. Rater 2 gives unexpected ratings compared to other raters, with about 7.82% unexpected ratings. Student 19 is judged biased by raters. Approximately 25.93% of the assessments given by all raters using all criteria for student 19 received unexpected responses. This information means that the rater needs clarification about assessing student 19 carrying out practicum activities. Item 9, related to communication, and item 8, related to the interpretation, are difficult for all raters to assess.

Table 7. Unexpected Responses

Cat.	Score	Exp.Score	Res.	Res. SE	Raters	Student	Items
3	3	3.90	-0.90	-3.20	Rater_2	Student 19	1 Moving
3	3	3.90	-0.90	-3.20	Rater_2	Student 20	1 Moving
3	3	3.90	-0.90	-3.00	rater_1	Student 19	1 Moving
3	3	3.90	-0.90	-3.00	rater_1	Student 20	1 Moving
3	3	3.90	-0.90	-2.80	rater_3	Student 19	1 Moving
3	3	3.90	-0.90	-2.80	rater_3	Student 20	1 Moving
1	1	2.50	-1.50	-2.70	rater_2	Student 6	8 Interpretation
2	2	3.10	-1.10	-2.60	rater_1	Student 12	8 Interpretation
4	4	2.80	1.20	2.60	rater_1	Student 19	7 Cleanliness
4	4	2.80	1.20	2.60	rater_2	Student 19	7 Cleanliness
1	1	2.40	-1.40	-2.60	rater_3	Student 14	9 Communication
4	4	2.80	1.20	2.60	rater_3	Student 19	7 Cleanliness
4	4	3.00	1.00	2.50	rater_1	Student 12	4 Assembling
1	1	2.40	-1.40	-2.50	rater_2	Student 1	8 Interpretation
1	1	2.30	-1.30	-2.40	rater_1	Student 1	8 Interpretation
4	4	3.00	1.00	2.40	rater_2	Student 12	4 Assembling
3	3	1.70	1.30	2.40	rater_3	Student 26	9 Communication
3	3	1.70	1.30	2.30	rater_1	Student 17	9 Communication
3	3	1.70	1.30	2.30	rater_1	Student 26	9 Communication
1	1	2.30	-1.30	-2.30	rater_2	Student 7	8 Interpretation
3	3	1.80	1.20	2.30	rater_2	Student 17	9 Communication
3	3	3.80	-0.80	-2.30	Rater_2	Student 18	3 Work Safety
3	3	1.80	1.20	2.30	Rater_2	Student 26	9 Communication
2	2	2.90	-0.90	-2.30	Rater_3	Student 12	9 Communication
1	1	2.20	-1.20	-2.20	Rater_1	Student 19	9 Communication
1	1	2.20	-1.20	-2.20	Rater_2	Student 2	8 Interpretation
3	3	1.80	1.20	2.20	Rater_2	Student 21	4 Assembling
3	3	1.90	1.10	2.20	Rater_3	Student 2	4 Assembling
3	3	1.80	1.20	2.20	Rater_3	Student 10	8 Interpretation
3	3	1.90	1.10	2.10	rater_1	Student 10	8 Interpretation
3	3	1.90	1.10	2.10	rater_1	Student 15	9 Communication
3	3	3.80	-0.80	-2.10	rater_1	Student 18	3 Work Safety
3	3	1.90	1.10	2.10	rater_2	Student 2	4 Assembling
3	3	1.90	1.10	2.10	Rater_2	Student 9	9 Communication
3	3	1.90	1.10	2.10	Rater_2	Student 10	8 Interpretation
1	1	2.10	-1.10	-2.10	Rater_2	Student 21	8 Interpretation
1	1	2.10	-1.10	-2.10	Rater_2	Student 23	8 Interpretation
1	1	2.10	-1.10	-2.10	rater_3	Student 20	4 Assembling
1	1	2.10	-1.10	-2.00	rater_1	Student 1	9 Communication
1	1	2.10	-1.10	-2.00	rater_2	Student 11	8 Interpretation
3	3	1.90	1.10	2.00	rater_2	Student 15	9 Communication
Top 3 Unexpected Responses		Overvalue (2.61%)		Rater_2 (7.82%)		Student 19 (25.93%)	9 Communication (14.81%)
		Undervalue (3.02%)		Rater_1 (5.35%)		Student 20 (14.81%)	8 Interpretation (14.81%)
		-		Rater_3 (3.70%)		Student 12 (14.81%)	4 Assembling (7.41%)

Note. Cat. = Category, Exp. Score = Expected Score, Res. =Residual, Res. SE =Residual Standard Error

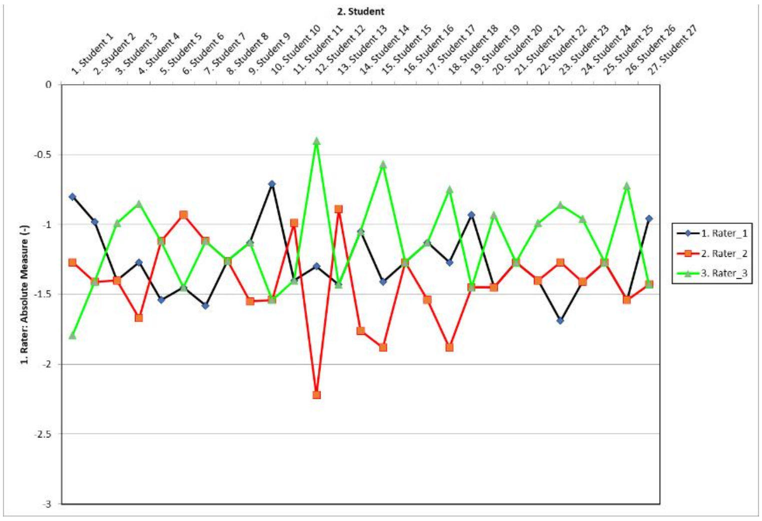


Figure 2. Bias/ Interaction Rater and Student-Based Absolute on Rater Measure

The output of Facets does not show any significant bias between raters and students. Figure 2 shows the absolute rating of the rater for all students. Table 3 regarding rater information shows that rater 3 tends to be more severe while rater 2 tends to be more flexible. Figure 2 shows that there are 5 cases of assessment bias by the rater, namely: student 1, student 6, student 9, student 11, and student 13, apparently rated more linear by rater 3 than rater 2, which should make rater 3 judge more server.

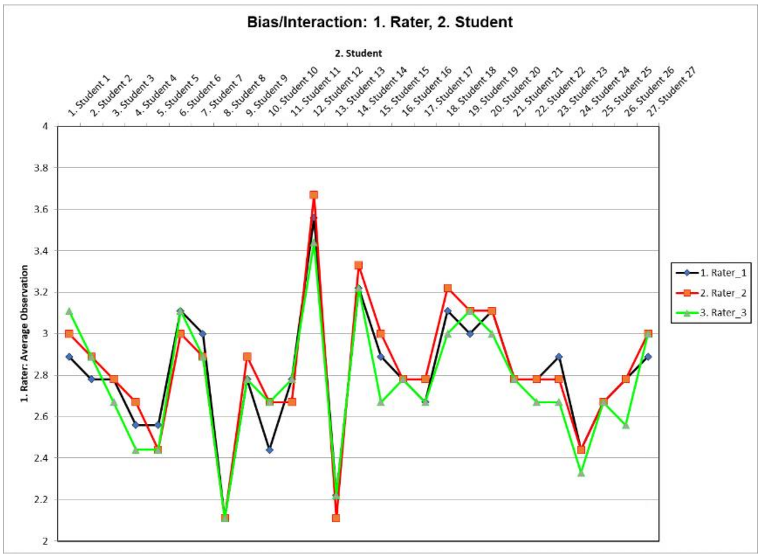


Figure 3. Bias/ Interaction Between Rater and Student Based on Average Rater Observation.

Figure 3 shows the average rater's assessment for all students, which provides information that all raters rate students consistently. This information is also an illustration of the Exact agreement rater, which reached 86.6% (see

Table *4*). No rater gives an extreme rating (extremely high or extremely low). Under certain conditions, an Exact agreement close to 100% and far from the expected agreement can be interpreted as the renters working together so that the ratings are similar.

Rating Scale Functionality

Table 8 shows that all categories have 9% to 35% responses. This data shows that the distribution of responses given to all categories of student practicum performance assessments is not uniform. The mean of the measurement shows the average for each category. The average measurement increases monotonically based on the student's ability level, which supports the assumption that the high category is given to students with high abilities. In contrast, the low category is given to students with low abilities. Meanwhile, the average measurement was similar to what the model expected. These data are supported by the model fit showing that all OMS indices are close to 1. Category 1 =1, category 2 = .8 and categories 3 and 4 = 1. In addition, the threshold value of the Rasch-Andrich limit shows an increasing monotonic trend, namely: -3.74 between categories 1 and 2, -.41 between categories 2 and 3, and 4.15 between categories 3 and 4. This information indicates that the categories used to assess student practicum performance are functioning well.

Table 8. Scales Functionality of the Scoring Rubrics.

Scoring Categories	Total	Percent	Average measures	Expected measures	MNSQ Outfits	Rasch- Andrich Thresholds
1	63	9%	-3.09	-3.25	1.1	-
2	218	30%	-1.81	-1.72	.8	-3.74
3	252	35%	1.97	1.98	1.0	-.41
4	196	27%	5.16	5.11	1.0	4.15

Note. MNSQ = mean square

Discussion

The analysis results show that in terms of content, the criteria/items used to assess practicum activities have a level of validity that meets the criteria. This criterion consists of the practicum process and product assessment, essential components in practicum activities. High school students who will enter university, especially in the chemistry programme, are expected to have the ability to conduct qualified practicums (Hensiek et al., 2016; Zengele & Alemayehu, 2016).

Process and product assessment are to provide a comprehensive picture of students' abilities. These two components will provide teachers with information about the ability to work in the classroom of students (Porter et al., 2017). Student performance assessment is an alternative for teachers to assess students' abilities for cognitive and psychomotor aspects of students (Chen et al., 2013; Cougar Hall & West, 2011). Student performance in the laboratory cannot be revealed directly only through paper and pencil tests but requires detailed completion of the minimum aspects that students carry out practicum (Harsh, 2016).

This study identified nine assessment criteria to assess students' practicum performance. These criteria include the ability to move, discipline, work safety, assembly, attitude, thoroughness, cleanliness, interpretation, and communication. This case is based on the criteria suggested by several previous studies, which state that the criteria for practicum activities include process and product capabilities (Irwanto et al., 2018; Royal Society of Chemistry, n.d.; Tahya et al., 2022)

The results of the MFRM analysis show the model's suitability for all aspects of the assessment category, students' abilities, raters, and the criteria or items used. All aspects are in the range of .5 – 1.5 intervals. Although there are five students and two criteria that are still outside the acceptance interval, it can be overcome by conditioning and revision in practice. The rater component and the rating category have a high degree of match between the other components. The rater component assesses according to the expected model, possibly influenced by the direction given before using the scoring rubric, the clarity of the use of the rubric, and the predictors of each category that are well understood by the rater. Unexpected assessment events may be influenced by items and students that do not fit the model, which can be seen in the model. the bias pattern in Table 7 shows that the raters understand the criteria and predictors developed assessment rubric well. According to Brennan (2010) and Yan (2014), if the raters understand well all the assessment criteria and predictors, then the understanding between them will be high even though they work independently.

The separation index shows that the rater can consistently differentiate students' abilities into three groups: high, medium, and low. This condition shows that the distribution of student abilities is relatively wide so that the reliability of measurement is high and the measurement error is relatively low (Eckes, 2015). This condition means that the assessment given to students by the rater will provide consistent results. In addition, the assessment criteria/items used in this study will maintain their level of difficulty when used in samples with the same student population in this study. The studies conducted show that there is no significant bias between students and raters. This study is one result of good measurement quality. Increasing the reliability of facets can be done by eliminating bias. This effort can increase the range of students' abilities to be more comprehensive, rater understanding also increases, and item differentiability increases (Eckes, 2015; Uto, 2021).

In general, the effectively developed measuring instrument can be used at intervals of (-4.04, 4.26) logit. Students' ability from the interval (-2.86, 4.26) logit entered at this measurement interval. The developed item has a difficulty index from (-4.04, 3.87) logit. The rater has a severity level in the interval (-1.39, -1.13) logit. Based on this information, with the severity index owned by the raters, the probability of students graduating is 23 people. Criteria for discipline (2), work safety (3), attitude (5), and movement (1) are items of criteria that are easy for all students to do. The cleanliness (7) and accuracy (6) criteria can only be done well by two students, while 25 other students may fail on this item. Although the criteria for communication (9), assembling (4) and interpretation (8) are difficult items for almost all students. This condition shows that of the nine criteria, the last three criteria need to be considered by teachers in chemistry practicum activities.

The communication criteria are part of the product assessment, and the communication criteria show the ability of students to convey their ideas both in writing in reports and explain the activities carried out during practicum activities (Giammatteo & Obaya, 2018). Assembling criteria is the ability of students to assemble practical tools according to the instructions in the module proportionally and adequately (Asmorowati et al., 2021). Meanwhile, the interpretation criteria show the student's ability to explain the practicum results obtained based on the relevant theory (Turiman et al., 2012).

Communication ability is an essential criterion for students. Students' ability to convey ideas orally or in writing in reports requires particular cognitive abilities and skills, so only a few students can achieve these skills, especially in practical activities (Hensiek et al., 2016). students generally can carry out practical activities from beginning to end quickly, but communicating what is being carried out and the results obtained are challenging criteria. Students' communication skills can be trained by allowing them to discuss in groups. In addition, the teacher can give assignments to students to write down ideas on specific topics, and then the teacher gives feedback and scaffolding to students (Reigosa & Jiménez-Aleixandre, 2007). The teacher can also occasionally allow students to express their aspirations so that they are more confident in communicating. Several studies (Montgomery et al., 2022; Reynders et al., 2019; Skagen et al., 2018) have proven to improve student's communication skills, especially in practical activities.

The components of assembling practicum tools show very difficult criteria based on item difficulty. This shows that almost all students need help to assemble chemistry practicum tools correctly and proportionally (Neamah, 2020; J. S. Wright et al., 2018). This criterion needs to be a concern for teachers to pay more attention to the ability of students to assemble tools so that practicum activities can run by the objectives of the practicum. Several studies have shown that students are afraid of assembling tools for fear of breaking or breaking tools during practicum (D'Souza et al., 2017; Straut & Nelson, 2020); they do not understand well the procedure for assembling practical tools (Chairam et al., 2015; Reigosa & Jiménez-Aleixandre, 2007), and some teachers do not allow students to assemble their tools so that they are not proficient (Cheung, 2008, 2011). These reasons are thought to cause students to have very low abilities on the item criteria for assembling tools. To solve this problem, teachers should use virtual media or videos to gather practical tools (Harwood et al., 2020; Hennah & Seery, 2017). Additionally, teachers can also use Web-based teaching aids to increase their confidence in assembling tools and materials without fear of damaging the tools (Hanifah et al., 2021). The most important thing is that the teacher is positioned to accompany, direct, and guide students in assembling tools before practicum activities.

The ability to interpret practicum results because it is helpful for students to describe and explain the meaning of data, and thus students will be easy to understand the data obtained (Lunardi et al., 2021). According to Subali et al. (2015) emphasise that the ability to interpret data is related to understanding the concepts they have. Interpretation is not just reading but emphasizing understanding concepts and interpreting them based on related theories (Sa'adah et al., 2020). Interpretive ability is part of higher-order thinking skills, but teachers have yet to direct students in higher-order thinking processes (Lunardi et al., 2021; Sa'adah & Sigit, 2018). In addition, interpretation skills are considered difficult for students and students at school, and this is because these abilities are related to spatial (spatial) abilities (Subali et al., 2015), logical abilities and mathematical abilities. The solution teachers can do to improve students' interpretation skills is to direct students to use high-level abilities in learning (R. A. Hunter & Kovarik, 2022; Lichti et al., 2021). In addition, several studies have shown the application of science. The approach writing heuristic (SWH) can significantly improve students' interpretation ability (Rudd et al., 2007; Sa'adah et al., 2020).

Based on the results obtained from studies and theoretical studies, exciting findings are obtained that the ability to interpret data is a criterion that is interrelated with communication criteria and the ability to assemble tools. These two components are interrelated with each other. Students who correctly assemble tools will produce valid and correct data. Eventually, the resulting interpretation will also be wrong. As a result, wrong interpretation of the data causes the quality of the information produced to be less valid. This case is also the other way around. This finding is evidence that all performance criteria in practicum activities are related, so all criteria must be considered and improved.

Conclusion

A chemistry practicum is a series of activities requiring direct observation to assess students' abilities. One of the commonly used assessments is the use of an assessment rubric. The quality of the instruments developed shows that all facets fit with the model and have good reliability. Homogeneous rater conditions indicate rater understanding of using the scoring rubric using the nine criteria developed to assess students' abilities with precision. The categories of each criterion/item all work fine. Unexpected responses were also relatively few, namely only 6% of all responses. The rubric component developed in this study includes nine criteria for assessing practicum processes and products. Components that require attention in practicum activities are the ability to assemble tools, communication skills, interpretation abilities, cleanliness, and student accuracy. Based on this finding, these five criteria must be of particular concern to teachers. Students' ability to carry out practicums will be even better if these five components have been well mastered, so teachers need to strive for the best way so that their students master these criteria.

Recommendations

Finally, one of the strengths of MFRM is providing a specific and detailed description of the facets so that it can provide comprehensive and valuable information in the development of measuring instruments. Regarding the context of the development of measuring instruments in this study, of the nine items developed, it can be seen that the item components need to be the focus of the findings of this study, given the consistent rater components and relatively well-distributed student abilities. It is hoped that future research will examine other components, such as rater and student, by increasing the number and variability of each component. Recommendations for future research regarding assessments involving raters, it is necessary to consider various rates so that they can reveal several components that are sources of bias, for example, gender, ethnicity, and race of raters. MFRM analysis is a recommendation for future researchers interested in finding information from all aspects of measurement individually and holistically.

Limitations

This study has several limitations, first is the relatively small number of raters, only three people, so only a little information is found for this facet. Many raters can show several levels of severity raters which can be used as a reference when we determine research objectives. For example, a rater with high severity is needed during tight selection activities, while a standard test requires a moderate severity level. Second, the criteria used still need to be reduced to several items that can specifically be used as a reference to assess students' abilities in practicum activities, and the authors realise that they still need to add other criteria that can be used as references/standards to assess students' abilities in practical activities.

Acknowledgements

The authors would like to thank for Universitas Islam Negeri Maulana Malik Ibrahim Malang all its facilities to support this research.

Authorship Contribution Statement

Elvira: Conceptualization, design, and writing. Retnawati: Review and supervision. Rohaeti: Supervision and final approval. Sainuddin: Statistical analysis, reviewing, drafting the manuscript, and admin.

References

Adams, C. J. (2020). A constructively aligned first-year laboratory course. *Journal of Chemical Education*, *97*(7), 1863–1873. https://doi.org/10.1021/acs.jchemed.0c00166

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, *45*(1), 131–141. https://doi.org/10.1177/0013164485451012

Almarshoud, A. F. (2011). Developing a rubric-based framework for measuring the ABET outcomes achieved by students of electric machinery courses. *International Journal of Engineering Education*, *27*(4), 859–866.

Aryadoust, V. (2016). Gender and academic major bias in peer assessment of oral presentations. *Language Assessment Quarterly*, *13*(1), 1–24. https://doi.org/10.1080/15434303.2015.1133626

Asmorowati, D., Wardani, S., & Mahatmanti, F. (2021). Analysis of student science process skills in the practicum of physical chemistry based on linguistic and interpersonal intelligence. *International Journal of Active Learning*, *6*(1), 34–40. https://www.learntechlib.org/p/218989/

Basturk, 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

Bennett, R. E., Deane, P., & van Rijn, P. W. (2016). From cognitive-domain theory to assessment practice. *Educational Psychologist*, *51*(1), 82–107. https://doi.org/10.1080/00461520.2016.1141683

Beyreli, L., & Ari, G. (2009). The use of analytic rubric in the assessment of writing performance-inter-rater concordance study. *Educational Sciences: Theory and Practice*, *9*(1), 105–125. https://hdl.handle.net/20.500.12451/6891

Bodner, G. M. (2015). Research on problem solving in chemistry. *Chemistry Education: Best Practices, Opportunities and Trends, January 2015*, 181–202. https://doi.org/10.1002/9783527679300.ch8

Bond, T., Yan, Z., & Heene, M. (2020). *Applying the rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge. https://doi.org/10.4324/9780429030499

Brennan, R. L. (2010). Generalizability theory and classical test theory. *Applied Measurement in Education*, *24*(1), 1–21. https://doi.org/10.1080/08957347.2011.532417

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Association for Supervision & Curriculum Development. http://bit.ly/3JkVojK

Capozzi, F., Laghi, L., & Belton, P. S. (Eds.). (2015). *Magnetic resonance in food science: Defining food by magnetic resonance*. The Royal Society of Chemistry. https://doi.org/10.1039/9781782622741

Chairam, S., Klahan, N., & Coll, R. K. (2015). Exploring secondary students' understanding of chemical kinetics through inquiry-based learning activities. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 937–956. https://doi.org/10.12973/eurasia.2015.1365a

Chen, H.-J., She, J.-L., Chou, C.-C., Tsai, Y.-M., & Chiu, M.-H. (2013). Development and application of a scoring rubric for evaluating students' experimental skills in organic chemistry: An instructional guide for teaching assistants. *Journal of Chemical Education*, *90*(10), 1296–1302. https://doi.org/10.1021/ed101111g

Cheung, D. (2008). Facilitating chemistry teachers to implement inquiry-based laboratory work. *International Journal of Science and Mathematics Education*, *6*, 107–130. https://doi.org/10.1007/s10763-007-9102-y

Cheung, D. (2011). Teacher beliefs about implementing guided-inquiry laboratory experiments for secondary school chemistry. *Journal of Chemical Education*, *88*(11), 1462–1468. https://doi.org/10.1021/ed1008409

Chukwuere, J. E. (2021). The comparisons between the use of analytic and holistic rubrics in information systems discipline. *Academia Letters*, Article 3579. https://doi.org/10.20935/al3579

D'Souza, M. J., Roeske, K. P., & Neff, L. S. (2017). Free inventory platform manages chemical risks, addresses chemical accountability, and measures cost-effectiveness. *International Journal of Advances in Science, Engineering and Technology*, *5*(3), 25–29. https://bit.ly/40v0HUC

Deviana, T., Hayat, B., & Suryadi, B. (2020). Validation of the social provision scale with indonesian student sample: A rasch model approach. *Indonesian Journal of Educational Assesment*, *3*(1), Article 1.

Eckes, T. (2015). *Introduction to many-facet rasch measurement: analyzing and evaluating rater-mediated assessments* (2nd ed.). Peter Lang Verlag. https://doi.org/10.3726/978-3-653-04844-5

Fisher, W. P., Jr. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transaction*, *21*(1), 1095. https://www.rasch.org/rmt/rmt211m.htm

Galti, A. M., Saidu, S., Yusuf, H., & Goni, A. A. (2018). Rating scale in writing assessment: Holistic vs. Analytical scales: A review. *International Journal of English Research*, *4*(6), 4–6.

Ghaemi, R. V., & Potvin, G. (2021). Hands-on education without the hands-on? An approach to online delivery of a senior lab course in chemical engineering while maintaining key learning outcomes. *Proceedings of the Canadian Engineering Education Association (CEEA)*, *2021*, 1-8. https://doi.org/10.24908/pceea.vi0.14834

Giammatteo, L., & Obaya, A. V. (2018). Assessing chemistry laboratory skills through a competency-based approach in high school chemistry course. *Science Education International*, *29*(2), 103–110. https://doi.org/10.33828/sei.v29.i2.5

Gürses, A., Çetinkaya, S., Doğar, Ç., & Şahin, E. (2015). Determination of levels of use of basic process skills of high school students. *Procedia - Social and Behavioral Sciences*, *191*, 644–650. https://doi.org/10.1016/j.sbspro.2015.04.243

Hager, P., Gonczi, A., & Athanasou, J. (1994). General issues about assessment of competence. *Assessment & Evaluation in Higher Education*, *19*(1), 3–16. https://doi.org/10.1080/0260293940190101

Hall, P. C., & West, J. H. (2011). Potential predictors of student teaching performance: Considering emotional intelligence. *Issues in Educational Research*, *21*(2), 145–161. http://www.iier.org.au/iier21/hall.html

Hanifah, S., Sari, & Irwansyah, F. S. (2021). Making of web-based chemical laboratory equipment and materials inventory application. *Seminar Nasional Tadris Kimiya 2020*, *2*, 97–110. http://bit.ly/3JNsVDB

Harmey, S., D'Agostino, J., & Rodgers, E. (2019). Developing an observational rubric of writing: Preliminary reliability and validity evidence. *Journal of Early Childhood Literacy*, *19*(3), 316–348. https://doi.org/10.1177/1468798417724862

Harsh, J. A. (2016). Designing performance-based measures to assess the scientific thinking skills of chemistry undergraduate researchers. *Chemistry Education Research and Practice*, *17*(4), 808–817. https://doi.org/10.1039/c6rp00057f

Harwood, C. J., Hewett, S., & Towns, M. H. (2020). Rubrics for assessing hands-on laboratory skills. *Journal of Chemical Education*, *97*(7), 2033–2035. https://doi.org/10.1021/acs.jchemed.0c00200

He, T.-H., Gou, W. J., Chien, Y.-C., Chen, I.-S. J., & Chang, S.-M. (2013). Multi-faceted Rasch measurement and bias patterns in EFL writing performance assessment. *Psychological Reports*, *112*(2), 469–485. https://doi.org/10.2466/03.11.PR0.112.2.469-485

Hennah, N., & Seery, M. K. (2017). Using digital badges for developing high school chemistry laboratory skills. *Journal of Chemical Education*, *94*(7), 844–848. https://doi.org/10.1021/acs.jchemed.7b00175

Hensiek, S., DeKorver, B. K., Harwood, C. J., Fish, J., O'Shea, K., & Towns, M. (2016). Improving and assessing student hands-on laboratory skills through digital badging. *Journal of Chemical Education*, *93*(11), 1847–1854. https://doi.org/10.1021/acs.jchemed.6b00234

Hlukhaniuk, V., Solovei, V., Tsvilyk, S., & Shymkova, I. (2020). STEMA education as a benchmark for innovative training of future teachers of labour training and technology. *Society. Integration. Education. Proceedings of the International Scientific Conference*, *1*, 211–221. https://doi.org/10.17770/sie2020vol1.5000

Hunter, C., Mccosh, R., & Wilkins, H. (2003). Integrating learning and assessment in laboratory work. *Chemistry Education Research and Practice*, *4*(1), 67–75. https://doi.org/10.1039/b2rp90038f

Hunter, R. A., & Kovarik, M. L. (2022). Leveraging the analytical chemistry primary literature for authentic, integrated content knowledge and process skill development. *Journal of Chemical Education*, *99*(3), 1238–1245. https://doi.org/10.1021/acs.jchemed.1c00920

Irwanto, Rohaeti, E., & Prodjosantoso, A. K. (2018). The investigation of university students' science process skills and chemistry attitudes at the laboratory course. *Asia-Pacific Forum on Science Learning and Teaching*, *19*(2), Article 07. http://bit.ly/3FzomeL

Janssen, G., Meier, V., & Trace, J. (2015). Building a better rubric: Mixed methods rubric revision. *Assessing Writing*, *26*, 51–66. https://doi.org/10.1016/j.asw.2015.07.002

Johnson, E. S., Zheng, Y., Crawford, A. R., & Moylan, L. A. (2019). Developing an explicit instruction special education teacher observation rubric. *Journal of Special Education*, *53*(1), 28–40. https://doi.org/10.1177/0022466918796224

Lichti, D., Mosley, P., & Callis-Duehl, K. (2021). Learning from the trees: Using project budburst to enhance data literacy and scientific writing skills in an introductory biology laboratory during remote learning. *Citizen Science: Theory and Practice*, *6*(1), Article 32. https://doi.org/10.5334/CSTP.432

Linacre, J. M. (1994a). *FACET: Rasch Model* (2nd ed.). Mesa Press.

Linacre, J. M. (1994b). *Many-facet: Rasch measurement* (2nd ed.). Mesa Press.

Linacre, J. M. (2002). Review of reviews of Bond & Fox (2001). *Rasch Measurement Transactions*, *16*(2), 871–882. https://www.rasch.org/rmt/rmt162.pdf

Linacre, J. M. (2018). *A user guide to FACETS Rasch-model computer programs*. Winsteps. https://bit.ly/3AEBqfQ

Lunardi, C. N., Gomes, A. J., Rocha, F. S., De Tommaso, J., & Patience, G. S. (2021). Experimental methods in chemical engineering: Zeta potential. *Canadian Journal of Chemical Engineering*, *99*(3), 627–639. https://doi.org/10.1002/cjce.23914

Maknun, D. (2015). Evaluasi keterampilan laboratorium mahasiswa menggunakan asesmen kegiatan laboratorium berbasis kompetensi pada pelaksanaan praktek pengalaman lapangan (PPL) [Evaluation of students' laboratory skills using competency-based laboratory activity assessment during the implementation of field experience practice (FEP)]. *Jurnal Tarbiyah*, *22*(1), 21–47. https://bit.ly/40iDjJT

Mistry, N., & Gorman, S. G. (2020). What laboratory skills do students think they possess at the start of University? *Chemistry Education Research and Practice*, *21*(3), 823–838. https://doi.org/10.1039/c9rp00104b

Mitchell, A. A. (2006). Review of the book introduction to rubrics: An assessment tool to save grading time, convey effective feedback and promote student learning. *Journal of College Student Development*, *47*(3), 352–355. https://doi.org/10.1353/csd.2006.0033

Montgomery, T. D., Buchbinder, J. R., Gawalt, E. S., Iuliucci, R. J., Koch, A. S., Kotsikorou, E., Lackey, P. E., Lim, M. S., Rohde, J. J., Rupprecht, A. J., Srnec, M. N., Vernier, B., & Evanseck, J. D. (2022). The scientific method as a scaffold to enhance communication skills in chemistry. *Journal of Chemical Education*, *99*(6), 2338–2350. https://doi.org/10.1021/acs.jchemed.2c00113

Morgan, G. B., Zhu, M., Johnson, R. L., & Hodge, K. J. (2014). Interrater reliability estimators commonly used in scoring language assessments: A Monte Carlo investigation of estimator accuracy. *Language Assessment Quarterly*, *11*(3), 304–324. https://doi.org/10.1080/15434303.2014.937486

Neamah, W. Q. (2020). Academic laboratory skills for chemistry students at the college of education for pure sciences - Ibn Al Haitham. *Journal of Xi'an University of Architecture & Technology*, *XII*(III), 1531–1554.

Ng, S. B. (2019). *Exploring STEM competences for the 21st century* (C. Gallagher, L. Ji, & T. Kiyomi (Eds.)). UNESCO International Bureau of Education. https://bit.ly/40dMwmE

Orgill, M., York, S., & MacKellar, J. (2019). Introduction to systems thinking for the chemistry education community. *Journal of Chemical Education*, *96*(12), 2720–2729. https://doi.org/10.1021/acs.jchemed.9b00169

Porter, A. L., Barnett, S. G., & Gallimore, C. E. (2017). Development of a holistic assessment plan to evaluate a four-semester laboratory course series. *American Journal of Pharmaceutical Education*, *81*(2), Article 33. https://doi.org/10.5688/ajpe81233

Pusca, D., Bowers, R. J., & Northwood, D. O. (2017). Hands-on experiences in engineering classes: The need, the implementation and the results. *World Transactions on Engineering and Technology Education*, *15*(1), 12–18. https://bit.ly/3JCLPg0

Reigosa, C., & Jiménez-Aleixandre, M.-P. (2007). Scaffolded problem-solving in the physics and chemistry laboratory: Difficulties hindering students' assumption of responsibility. *International Journal of Science Education*, *29*(3), 307–329. https://doi.org/10.1080/09500690600702454

Reynders, G., Suh, E., Cole, R. S., & Sansom, R. L. (2019). Developing student process skills in a general chemistry laboratory. *Journal of Chemical Education*, *96*(10), 2109–2119. https://doi.org/10.1021/acs.jchemed.9b00441

Royal Society of Chemistry. (n.d.). *Curriculum support*. https://rsc.li/3ng0y9I

Rudd, J. A., Greenbowe, T. J., & Hand, B. M. (2007). Using the science writing heuristic to improve students' understanding of general equilibrium. *Journal of Chemical Education*, *84*(12), 2007–2011. https://doi.org/10.1021/ed084p2007

Sa'adah, E. N. L., & Sigit, D. (2018). Pengembangan instrumen penilaian sikap dan keterampilan psikomotorik pada materi elektrokimia [Development of attitudes and psychomotor skills assessment instruments in electrochemical materials]. *Teori, Penelitian, Dan Pengembangan*, *3*(8), 1023-1026. https://bit.ly/3oMT0LW

Sa'adah, N., Langitasari, I., & Wijayanti, I. E. (2020). Implementasi pendekatan science writing heuristic pada laporan praktikum berbasis multipel representasi terhadap kemampuan interpretasi [Implementation of the science writing heuristic approach to multiple representation-based practicum reports on interpretation. *Jurnal Inovasi Pendidikan IPA*, *6*(2), 195–208. https://doi.org/10.21831/jipi.v6i2.31078

Sainuddin, S., Subali, B., Jailani, & Elvira, M. (2022). The development and validation prospective mathematics teachers holistic assessment tools. *Ingénierie des Systèmes d'Information*, *27*(2), 171–184. https://doi.org/10.18280/isi.270201

Seery, M. K. (2020). Establishing the laboratory as the place to learn how to do chemistry. *Journal of Chemical Education*, *97*(6), 1511–1514. https://doi.org/10.1021/acs.jchemed.9b00764

Seery, M. K., Agustian, H. Y., Doidge, E. D., Kucharski, M. M., O'Connor, H. M., & Price, A. (2017). Developing laboratory skills by incorporating peer-review and digital badges. *Chemistry Education Research and Practice*, *18*(3), 403–419. https://doi.org/10.1039/c7rp00003k

Skagen, D., McCollum, B., Morsch, L., & Shokoples, B. (2018). Developing communication confidence and professional identity in chemistry through international online collaborative learning. *Chemistry Education Research and Practice*, *19*(2), 567–582. https://doi.org/10.1039/c7rp00220c

Straut, C. M., & Nelson, A. (2020). Improving chemical security with material control and accountability and inventory management. *Journal of Chemical Education*, *97*(7), 1809–1814. https://doi.org/10.1021/acs.jchemed.9b00844

Subali, B., Rusdiana, D., Firman, H., & Kaniawati, I. (2015). Analisis kemampuan interpretasi grafik kinematika pada mahasiswa calon guru fisika [Analysis of kinematics graph interpretation ability in prospective physics teacher students]. *Prosiding Simposium Nasional Inovasi Dan Pembelajaran Sains 2015*, *3*(1), 269–272.

https://bit.ly/3JLjNz7

Tahya, D., Dahoklory, F. S., & Dahoklory, S. R. (2022). The development of local wisdom-based chemistry modules to improve students' science process skills. *Jurnal Penelitian Pendidikan IPA*, *8*(2), 731–739. https://doi.org/10.29303/jppipa.v8i2.1424

Turiman, P., Omar, J., Daud, A. M., & Osman, K. (2012). Fostering the 21st century skills through scientific literacy and science process skills. *Procedia - Social and Behavioral Sciences*, *59*, 110–116. https://doi.org/10.1016/j.sbspro.2012.09.253

Ural, E. (2016). The effect of guided-inquiry laboratory experiments on science education students' chemistry laboratory attitudes, anxiety and achievement. *Journal of Education and Training Studies*, *4*(4), 217–227. https://doi.org/10.11114/jets.v4i4.1395

Uto, M. (2021). A multidimensional generalized many-facet Rasch model for rubric-based performance assessment. *Behaviormetrika*, *48*, 425–457. https://doi.org/10.1007/s41237-021-00144-w

Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing*, *15*(2), 263–287. https://doi.org/10.1177/026553229801500205

Wesolowski, B. C. (2012). Understanding and developing rubrics for music performance assessment. *Music Educators Journal*, *98*(3), 36–42. https://doi.org/10.1177/0027432111432524

Wesolowski, B. C., Amend, R. M., Barnstead, T. S., Edwards, A. S., Everhart, M., Goins, Q. R., Grogan, R. J., Herceg, A. M., Jenkins, S. I., Johns, P. M., McCarver, C. J., Schaps, R. E., Sorrell, G. W., & Williams, J. D. (2017). The development of a secondary-level solo wind instrument performance rubric using the multifaceted rasch partial credit measurement model. *Journal of Research in Music Education*, *65*(1), 95–119. https://doi.org/10.1177/0022429417694873

Wright, B. D. (1996). Reliability and separation. *Rasch Measurement Transactions*, *9*(4), 472. https://www.rasch.org/rmt/rmt94n.htm

Wright, J. S., Read, D., Hughes, O., & Hyde, J. (2018). Tracking and assessing practical chemistry skills development: Practical skills portfolios. *New Directions in the Teaching of Physical Sciences*, *13*(1), Article 07. https://doi.org/10.29311/ndtps.v0i13.2905

Yamanishi, H., Ono, M., & Hijikata, Y. (2019). Developing a scoring rubric for L2 summary writing: A hybrid approach combining analytic and holistic assessment. *Language Testing in Asia*, *9*, Article 13. https://doi.org/10.1186/s40468-019-0087-6

Yan, X. (2014). An examination of rater performance on a local oral English proficiency test: A mixed-methods approach. *Language Testing*, *31*(4), 501–527. https://doi.org/10.1177/0265532214536171

Zengele, A. G., & Alemayehu, B. (2016). The status of secondary school science laboratory activities for quality education in Case of Wolaita Zone, Southern Ethiopia. *Journal of Education and Practice*, *7*(31), 1–11. https://bit.ly/3JPRDUN