JPPIPA 11(7) (2025)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

Assessing Students' Science Process Skills: Designing a Performance-Based Assessment Instrument for Saponification Lab Activities

Aztiannisa Zahra[1*], Nahadi[1], Wiwi Siswaningsih[1]

[1] Department of Chemistry Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

Received: May 11, 2025
Revised: June 18, 2025
Accepted: July 25, 2025
Published: July 31, 2025

Corresponding Author:
Aztiannisa Zahra
aztiannisa9@upi.edu

DOI: 10.29303/jppipa.v11i7.11757

Abstract: This study aims to develop a performance-based assessment instrument to assess students' Science Process Skills (SPS) for saponification laboratory activities. This research is an instrument development study, consisting of four main stages: planning, development, validation, and pilot testing. A total of twelve 11th grade students from a high school in Bandung participated in pilot testing phase to evaluate the instrument's reliability. This performance-based instrument covers 27 performance aspects aligned with six SPS indicators: predicting, planning, implementing, observing, communicating, and interpreting. This instrument demonstrates strong content validity, with a CVI of 0.99. Inter-rater reliability was analyzed using Kendall's W Coefficient. The results show variable agreement results: 2 aspects showed very strong agreement, 3 aspects showed strong agreement, and 11 aspects showed weak agreement. These findings suggest the need to refine several instrument items to enhance inter-rater reliability. The practicality of the instrumen, evaluated through assessors, resulted in a score of 3.35 on a scale of 4.00, indicating the instruments is very practical for further implementation. Overall, the final product of instrument is valid and practical for assessing students' SPS in the saponification lab activities, though further improvements are necessary to improve inter-rater agreement.

Keywords: Assessment; Performance-based instrument; Saponification Lab Activities; Science Process Skills

Introduction

Assessment is an integrated process in learning. It is the process of collecting, reporting, and utilizing information regarding the learning process and outcomes, as well as the competencies achieved by students (Nahadi & Firman, 2019; Schellekens et al., 2021). The information is obtained through various assessment instruments design to align with instructional objective (Nahadi & Firman, 2019). According to *Permendikbud* No.23 of 2016, the three aspects that need to be developed in the learning process are aspects of affective, knowledge, and psychomotor. However, assessments in chemistry learning in Indonesia are still largely focused on the cognitive domain (Ainillana & Louise, 2024; Sugrah et al., 2019). Ideally, assessment should reflect all three domains in a balanced manner (Lestari et al., 2025). This align with *Permendikbud* No.21 of 2022, article 5, which states that the selection and/or development of assessment instruments must consider students' characteristics and needs, also be based on the assessment plan outlined in the instructional design.

In the context of chemistry learning, the integration of cognitive, affective, and psychomotor domains is crucial and can be achieved effectively through practical activities because chemistry is an 'experimental science' (Nahadi & Firman, 2019). Practical activities enable

How to Cite:
Zahra, A., Nahadi, & Siswaningsih, W. (2025). Assessing Students' Science Process Skills: Designing a Performance-Based Assessment Instrument for Saponification Lab Activities. *Jurnal Penelitian Pendidikan IPA*, *11*(7), 702–712. https://doi.org/10.29303/jppipa.v11i7.11757

learners to directly apply the concept they have learned in class, reinforcing their understanding of chemical principles (Kriswantoro et al., 2021; Sugrah et al., 2019), and also engaging in laboratory activities fosters positives attitudes, enhances learning motivation, and contributes to a more meaningful and effective learning process (Okam & Zakari, 2017; Stephenson et al., 2020).

Furthermore, one important outcome of laboratory activities is the development of Science Process Skills (SPS), which are essential for students in the 21st century (Elfrida et al., 2021; Kusuma et al., 2025). SPS allows students to construct knowledge based on empirical evidence and apply it in real-life contexts (Sejati et al., 2020). Although SPS is recognized for its significance, research has revealed that its development through laboratory activities is not always optimal (Sudirman et al., 2023). According to Fiolida et al. (2021), certain SPS indicators remain underdeveloped in current laboratory practices, leading to inadequate learning outcomes. Research by Irwanto et al. (2018) & Rahayu (2020) support such concerns by revealing that students' SPS proficiency in chemistry classes are still relatively low and often unsatisfactory. These findings highlight the necessity of more focused and structured approaches for practical work to ensure that all aspects of SPS are effectively fostered in learning process.

A SPS assessment is required in order to determine the achievement of students' SPS (Sibic & Şesen, 2022; Sudirman et al., 2023). Students' SPS in practical activities can be assessed by performance-based assessment (Isnaini & Utami, 2020; Sudirman et al., 2023). Performance-based assessment is an assessment of process or skill, product, or both demonstrated by students, so that students are actively involved in hands-on and minds-on learning (Elfrida et al., 2021; Heydarnejad et al., 2022). Moreover, performance-based assessment is able to measure students' knowledge, reasoning, skills, products, and multiple intelligences (Sari et al., 2020). A preliminary study involving interviews with chemistry teacher revealed that performance-based assessment in schools are generally limited to cursory observations. This aligns with findings by Kusumaningtyas et al. (2018), which found that teachers in the field are frequently not apply authentic performance assessments. Instead, assessments are typically conducted through brief, non-formal observations which fall short of accurately capture the competencies that students should demonstrate.

Further evidence from Mudhakiyah et al. (2022) indicates that performance assessment instruments currently used in chemistry learning tend to lack clarity and specificity. For example, these instruments fail to distinguish between different topics or learning materials and are often to not equip the essential components such as detailed scoring criteria and performance descriptions. Furthermore, Hikmah et al. (2018) highlight to a poor conceptual and operational understanding of Science Process Skills (SPS) among chemistry teachers, which negatively impacts the accuracy and suitability of SPS assessments in the classroom. Given these challenges, there is obvious need for a well-designed performance-based assessment instrument in chemistry learning that can efficiently assess students' SPS using structured tasks and detailed scoring rubrics. Such assessments should be able to demonstrate the specific competencies targeted in practical chemistry activities. Moreover, to ensure their effectiveness, performances assessment instruments must meet essential quality standards, including validity, reliability, and practically (Oktavilani & Agustini, 2024; Popham, 2017).

The chemistry curriculum for Grade 12th includes the topic of macromolecules, particularly the sub-material on fats, as outlined in *Permendikbud* No.37 of 2018. This content is specifically addressed in basic standard competency (KD) 4.11, which requires students to analyze the information regarding the production and impact of macromolecular products. The topic of macromolecules cover classification such as polymers, carbohydrates, proteins, lipids and fats. However, the majority of this content is theoretical and descriptive, which often makes it seem abstract and challeging for students to understand (Saragih et al., 2021). A preliminary survey conducted at high school in Bandung shows that performance-based assessment was not being used to assess students' Science Process Skills (SPS) in the lipid sub-material, especially through the saponification lab activity. In facts this activity has great potential to improve students' comprehension the topic of lipid in macromolecules. Students may better comprehend the idea and build important scientific skills by conducting the saponification experiment, which turn lipid into soap (Putri et al., 2024).

These findings lead the research to explore the topic: "Assessing Students' Science Process Skills: Designing a Performance-Based Assessment Instrument for Saponification Lab Activities". This study aims to design performance-based assessment instrument to assess students' Science Process Skills (SPS) through saponification lab activity.

Method

This research is an instrument development study that applies the Development and Validation (D&V) phases, adapted from the phases proposed (Adams & Wieman, 2011). The research stages are shown in Figure 1.

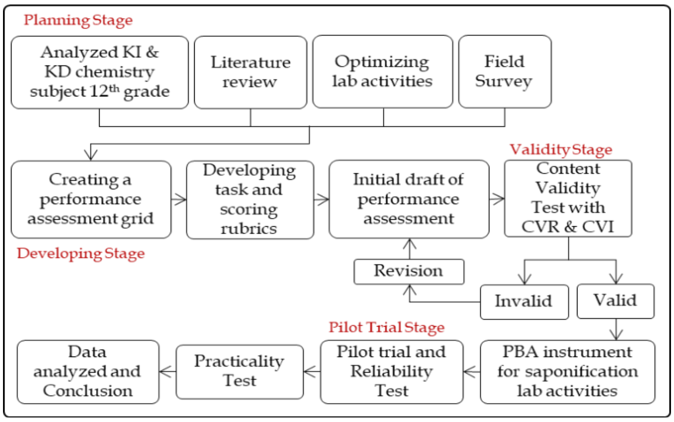


Figure 1. The stages of research process

The planning stage analyzed the core and standard competencies (KI and KD) of grade 12th chemistry curriculum (2013 revision), conducted a literature review, performed a field survey, and optimized the saponification activities. The development stage designed the performance assessment instrument, including creating an assessment grid, developed task, and prepared scoring rubrics. The validation stage focused on testing content validity to ensure the instrument's relevance and accuracy. Finally, in the pilot phase, a limited trial test was conducted, along with inter-rater reliability tests, to evaluate the consistency and usability of the instrument in the classroom.

This study involved 5 qualified validators, they are 2 high school chemistry teachers and 3 chemistry education lectures from Faculty of Mathematics and Natural Sciences Education, Universitas Pendidikan Indonesia (UPI). A minimum of 5 validators to guarantee adequate agreement (Almanasreh et al., 2019; Chong et al., 2021). In addition, 5 assessors (observers), who participated in this study, are chemistry education students from Faculty of Mathematics and Natural Sciences Education, Universitas Pendidikan Indonesia (UPI). The limited trial of the performance-based instrument product involved 12 of 11th graders who had previously studied the lipid sub-material in the macromolecule topic.

The instruments used in this study included an interview-guidelines, a validation sheet, an observation sheet for the limited trial and reliability testing, a student worksheet for collecting students' responses, and a questionnaire to evaluate the practicality of the performance-based instrument. The collected data were then analyzed. Content Validity Ratio (CVR) and Content Validity Index (CVI) methods were used to calculate the content validity test (Lawshe, 1975). The CVR value is obtained using the following Formula 1.

$$\mathrm{CVR} = \frac{n_e - \frac{N}{2}}{\frac{N}{2}} \quad (1)$$

As for n_e: number of validators stating valid and N: total number of validators. The minimum CVR value is 0.99 for five validators. Furthermore, The CVI value is obtained using the following Formula 2.

$$\mathrm{CVI} = \frac{\mathrm{CVR_t}}{\text{Total instrument items}} \quad (2)$$

As for CVR_t: CVR total.

The scores gathered from assessors in limited trial stage were then analyzed using an inter-rater reliability test in order to determine the rater's level of agreement. The test used the Kendall's W coefficient, which is suitable for ordinal data rated by numerous assessors (Gisev et al., 2013). The Kendall's W value was calculated using IBM SPSS 27, and its interpretation followed guidelines from (Hajghasem et al., 2022).

Table 1. Kendall's W Coefficient Interpretation

Kendall's W Coefficient	Agreement Interpretation
>0.700	Very Strong
0.510 – 0.700	Strong
0.310 – 0.500	Medium
0.110 – 0.300	Weak
<0.110	Very Weak

After the limited trial was conducted involving 5 assessors, a practicality test was carried out by distributing questionnaires to assessors who had used the performance-based assessment instrument (Rosidin et al., 2023). To determine the level of practicality, a following formula adapted form (Widoyoko, 2012).

$$\textbf{Practicality Score} = \frac{\text{Total score obtained form assessors}}{\text{Total number of assessors} \times \text{Number of Questionnaire Items}}$$

Following are the categories of practicality score as shown in Table 2.

Table 2. Practicality Categories

Practicality Score	Category
3.26 – 4.00	Very practical
2.51 – 3.25	Practical
1.76 – 2.50	Not practical
1.00 – 1.75	Very not practical

Result and Discussion

The findings of this study encompass the developmental stages of the performance-based assessment instrument and its quality.

Development Stages of Performance Assessment Instrument

The development of performance assessment instrument to assess SPS for saponification lab activities involved 4 main stages: planning, development, validation, and limited trial. Firstly, the planning stage.

The initial step in this stage involved analyzing the core and standard competencies (KI & KD) to identify relevant material for developing performance-based assessment instruments. According to *Permendikbud* No.37 of 2018, the standard competencies (KD) 4.11 was selected: "Analyzing the results of information searches regarding the manufacture and impact of a product derived from macromolecules." The macromolecule topic, particularly the subtopic on lipids, was selected because it is suitable for exploration through practical activities.

Furthermore, literature reviews revealed that SPS can be affectively developed through hands-on experiments (Elfrida et al., 2021). The implementation of SPS enables students to construct scientific concepts independently (Komisia et al., 2022). SPS assessment also fosters positive scientific attitudes such as curiosity and openness to new ideas (Hasanah et al., 2020). Moreover, SPS assessments help teachers in designing more targeted and effective learning strategis (Setiono & Astuti, 2021). Subsequently, interviews with teachers revealed that hands-on activity in lab for the lipid topic had not yet been implemented. Additionally, performance-based assessments were still limited to simple observations without structured instruments. Also, teachers acknowledged the need to performance-based assessment that capable more precisely to assess students' SPS. Concurrently, optimization of the saponification lab activities was conducted to identify appropriate tools, materials, and procedures. The procedure, which was adapted from Hasibuan et al. (2019) and OLABS, "Saponification-The Process of Making Soap", included steps such as testing coconut oil solubility, conducting saponification reaction, identifying soap using foam test, and analyzing emulsion properties.

Secondly, the development stage. This stage begins with the creation of an assessment grid to guide the development of performance-based instruments (Ainillana & Louise, 2024). Assessment grid ensures each competency is assessed proportionally (Sofyan et al., 2019). Indicators of competency achievement are developed from standard competency (KD) 4.11 utilizing Science Process Skills (SPS) indicators based on Nahadi & Firman (2019). These indicators are tailored to student's skills relevant to saponification lab activity, including observation, interpretation, prediction, implementation, planning, and communication. Each indicator is associated to minds-on and hands-on activities, which form the basis for designing the performance-based assessment instrument. Following the procedures outlined (Rasyidin & Mansur, 2009), performance tasks were developed by identifying the relevant knowledge and skills based on KD 4.11 and SPS in the lipids sub-topic, designing performance tasks, resulting in 27 tasks, and establishing a scoring guideline using a 1 – 4 rating scale. Therefore, there are 7 skill indicators, 6 SPS indicators, and 27 performance tasks.

Table 3. Some items from performance-based instrument grid

Skills Indicator	SPS Indicator	Performance Tasks
4.11.1 Predicting saponification reaction from coconut oil	Prediction	4.11.1.1 Construct problem statement concerning the saponification reaction of coconut oi
4.11.2 Designing saponification experiment from coconut oil	Experiment Planning	4.11.2.1 Re-arrange the saponification procedure from coconut oil according to the provided video

The initial draft of the performance assessment instrument was developed based on an assessment grid containing skill indicators, SPS indicators, performance tasks, and scoring rubrics. The scoring rubric is analytical rubric with 1 -4 scale that was modified from (Chowdhury, 2018). It consists of 27 performance items with specific criteria. A score of 1 indicates the lowest skill level and 4 indicates the highest. Each score is described based on the students' demonstrated skills. This performance instrument draft was designed to assess students' SPS in saponification lab activity.

The following section will then provide an explanation of the outcomes of the third and fourth stages, which are the validation stage and limited trial stage.

Performance-based Assessment Instrumen Quality

Performance Instrument's Content Validity

An instrument is considered valid it accurately measures it is intended to measure. Content validity refers to how well the instrument represents the learning material (Yusup, 2018). Content validity is assessed through expert judgment (Lawshe, 1975).

In this study, content validity was evaluated by expert validators (Viyanti et al., 2023), who assessed the alignment between skills indicators, SPS indicators, performance tasks, and scoring rubrics. Validators also provided suggestion to improve the instrument. These suggestions then served as the base for revisions, ensuring that performance-based instrument is valid in terms of content and able to affectively measure students' SPS for saponification lab activities. Feedback from the validators was analyzed using the CVR and CVI (Lawshe, 1975). The CVR results are shown in the Figure 2.

Table 4. An example of an item from initial draft of the performance-based instrument

SPS Indicator	Performance Tasks	Rubric
Prediction	4.11.1.1 Construct problem statement concerning the saponification reaction of coconut oil.	Rubric Indicators: 1.Construct the problem statement related to the saponification reaction. 2.Construct the problem statement based on phenomena provided in the worksheet. 3.Construct the problem statement about how soap work. Scoring Guidelines: 4: Problem statements meet all the 3 indicators accurately. 3: Problem statements meet the 2 indicators accurately. 2: Problem statement meets the 1 indicator accurately. 1: Problem statement does not align with any of the indicators.

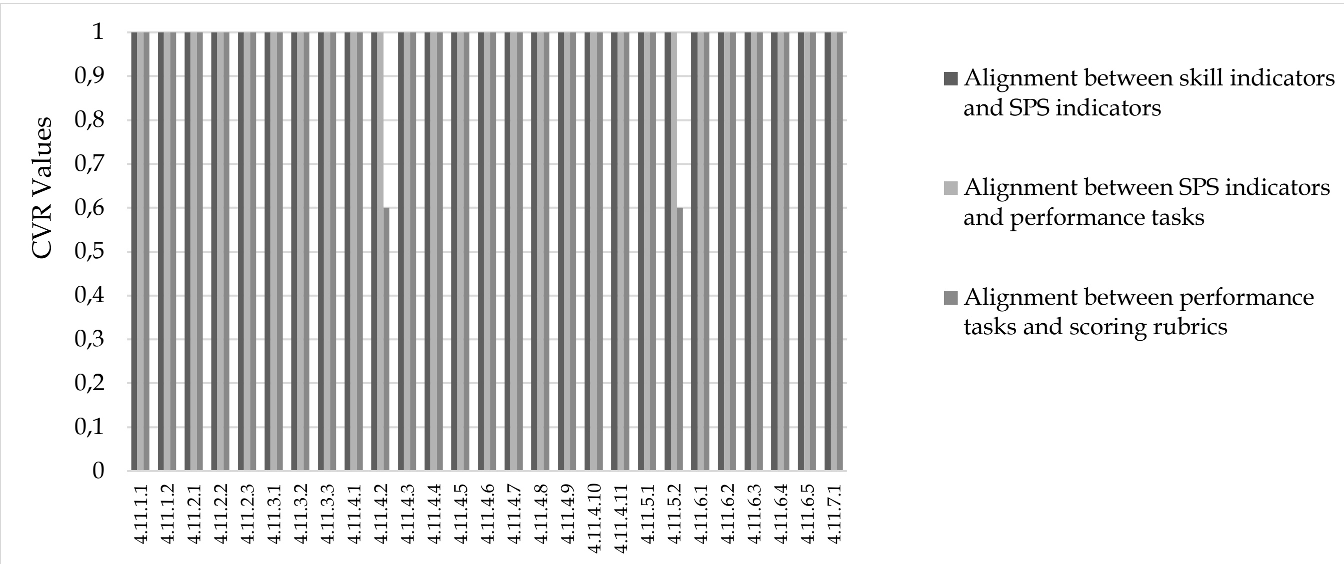


Figure 2. Graph of CVR Values for Instrument Items

Table 5. An example of task item revision

Before Revision	After Revision
Scoring Guidelines: 4: Measuring 5 mL of ethanol using a measuring cup correctly: the cup is placed on a flat surface, ethanol is poured along the inner wall, the lower meniscus is read at eye level, and the volume reading is accurate. 3: Measuring 5 mL of ethanol using a measuring cup correctly: the cup is placed on a flat surface, ethanol is poured along the inner wall, the lower meniscus is read at eye level, but the volume reading is not accurate. 2: Measuring 5 mL of ethanol using a measuring cup correctly: the cup is placed on a flat surface, ethanol is poured along the inner wall, but the lower meniscus is not read at eye level, and the volume reading is not accurate. 1: Measuring 5 mL of ethanol using a measuring cup correctly: the cup is placed on a flat surface, but ethanol is not poured along the inner wall, the lower meniscus is not read at eye level, and the volume reading is not accurate.	Indicators: 1. Places the measuring cylinder on a flat, level surface. 2. Pours ethanol carefully along the inner wall of the measuring cylinder without spilling. 3. Reads the lower meniscus at eye level. 4. Records the volume accurately based on the observed meniscus. Scoring Guidelines: 4: All 4 indicators are demonstrated correctly when measuring 5 mL of ethanol. 3: 3 indicators are demonstrated correctly. 2: 2 indicators are demonstrated correctly. 1: 1 indicator is demonstrated correctly.

The CVR values were compared with Lawshe (1975) minimum standard for 5 validators, which is 0.99. If the CVR is above 0.99, the item is valid, if below, it is not. All items related align between skill indicators, SPS indicators, and performance tasks with CVR values of 1, so they are considered valid. Out of 27 items, 25 had a CVR of 1, while 2 items scored 0.6 for alignment between performance tasks and the scoring rubrics (items 4.11.4.2 and 4.11.5.2). This score is below the minimum standard. However, those 2 items were still kept because the validators determined that they were relevant in terms of skill and performance alignment. According to (Lawshe, 1975), CVR results can be supplemented by additional approaches for determining whether to maintain particular items.

Furthermore, the CVI value is calculated as the average CVR values of the items that passed the minimum threshold. The following table shows the CVI values.

Table 6. CVI Result

Content Validity Category	CVI Value for Each Category	Overall CVI Value
Alignment between skill indicators and SPS indicators	1.00	
Alignment between SPS indicators and performance tasks	1.00	0.99
Alignment between performance tasks and scoring rubrics	0.95	

The overall CVI value of the performance-based assessment instrument to assess students' SPS for saponification lab activities is 0.99. This value exceeds the minimum standard suggested by Davis (1992), who recommended a CVI > 0.80. Since the result is above the threshold, the developed performance-based instrument is considered to have strong content validity. In many cases, using the overall CVI is also seen as more practical and efficient approach (Davis, 1992).

Performance Instrument's Inter-Rater Reliability

Reliability shows how consistently an instrument measures what it is supposed to (Livingston, 2018). The limited trial aimed to test reliability of the developed performance-based assessment instrument. The method used in this study was inter-rater reliability, which involves several assessors (observers) using the same instrument (Miller et al., 2009; Sullivan, 2011). Each rater gives a score based on students' performance (Miller et al., 2009). The results are then analyzed using Kendall's coefficient or concordance (W) to measure agreement among assessors (Gisev et al., 2013).

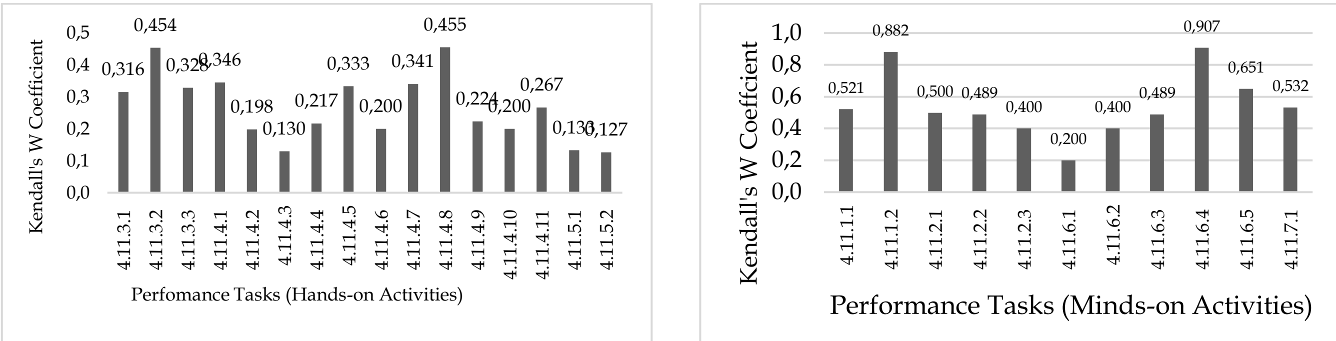
Figure 3. Left: A student performed the saponification lab activity; Right: An assessor observed the activities

A limited trial was conducted to test reliability of performance-based assessment instrument for assessing students' SPS. The trial involved 12 grades 11th students who had already learned about saponification in lipid sub-topic. The students were divided into 3 groups of 4 people. Firman (2013) outlines that one observer can observe 3 -6 students doing similar task at the same time. Each student's performance was assessed by assessors using the Google form-based instrument, which part 1 assessed hands-on activities and part 2 assessed minds-on activities. The obtained scores were analyzed using IBM SPSS 27. The Kendall's W coefficient was calculated for each aspect of the instrument to determine how consistent the agreements were. The results were then interpreted using from (Hajghasem et al., 2022) (Table 1).

The bar chart depicts the varying levels of agreement among assessors on 27 performance tasks. 2 aspects are classified as very strong agreement, 3 as strong agreement, 11 as moderate agreement, and another 11 as weak agreement. Amongst these, 16 aspects assess hands-on activities (physical tasks) and 11 aspects assess minds-on activities (knowledge tasks). The bar chart also depicts that the Kendall's W coefficients for hands-on activities are typically lower than that for minds-on activities. This shows that assessors had more difficult time making consistent judgment on physical tasks. Several factors may contribute to this, including limited visibility during observation (Wulan, 2018), scoring bias (Popham, 2017), and variances in how assessors apply the rubric's scoring scale (Wolf & Stevens, 2007).

Performance Instrument's Practicability

In addition to being valid and reliable, a good performance-based assessment instrument should also be practical (Murniati et al., 2023). A practical instrument is one that has clear instructions and guidelines, is simple to use, and is easy to score (Widoyoko, 2012). Assessors who had used the instrument were given the questionnaire to evaluate its practicality. A 4-point Likert Scale, with 4 (strongly agree), 3 (agree), 2 (disagree), 1 (strongly disagree), was employed in this questionnaire. This scale helps distinct viewpoints and avoids neutral answer (Widoyoko, 2012). The questionnaire was shared via Google Form during a limited trial. The practicality instrument evaluation covered 4 main aspects, which are feasibility (A), language (B), effectiveness (C), and weakness (D). These aspects were adapted from previous studies by (Siswaningsih et al., 2024).



(a) (b)

Figure 4. Graph of Kendall's W Coef: (a) Hands-on Activities Items; and (b) Minds-on Activities Items

Table 7. Instrument Practicality Questionnaire

Code	Aspect Statements
A1	SPS indicators are aligned with the skills being assessed.
A2	Performance aspects aligned with the descriptors in the rubric
A3	Performance aspects are consistent with SPS indicators
A4	The performance criteria in the rubric clearly reflect the performance aspects
A5	The scoring rubric reflect the defined performance criteria
A6	The differences between each score level in the rubric are clearly defined
B1	The rubric and performance aspects follow proper Indonesian writing rules
B2	The languages used for performance aspects is clear and understandable for assessors
B3	The language used in the rubric is clear and understandable for assessors
C1	Instructions for using the instrument are clear and easy to follow for assessors
C2	The 1-4 scoring scale on the rubric is simple to use
C3	The assessment rubric is easy to use
C4	The rubric's performance criteria are objective and fair
C5	The rubric accurately reflects students' performance on SPS indicators
C6	The indicators in the rubric are observable by assessors
C7	The performance aspects are arranged in a logical, observable sequence
C8	The rubric can be used to assess SPS during saponification lab activities
C9	The instrument can be used within the time allocated in the lesson plan
C10	The rubric supports easy processing of students' SPS
D1	The rubric is difficult to understand
D2	The score range in the rubric is too wide
D3	There are too many performance aspects

According to the bar chart form the practicality evaluation, both the feasibility aspect (A) and the language aspect (B) drew largely positives responses. The majority of respondents selected scores of 4 (strongly agree) and 3 (agree), indicating that instrument was considered practical and easy to understand. This finding aligns with the argument of Astuti (2012), who stated that an instrument's practicality is reflected through positive feedback from its assessors.

Regarding the effectiveness aspect (C), the majority of the responses were positive. However, there were a few exceptions. One respondent disagreed with statements C2, C9, C10, scoring a 2 on each. From statements of C2 and C10, this is in line with Wolf & Stevens (2007) explanation that a wider score range can make assessments slower and more complicated since it requires more time and judgment to decide on a score. However, the statement C9 disagreement in this case may be due to time needed to assess students' SPS.

Analytical rubric, the one used in this instrument, also tend to take longer to use compared to holistic rubric (Chowdhury, 2018). Despite these minor concerns, the overall responses suggest that the instrument is practical and largely well-received by assessors.

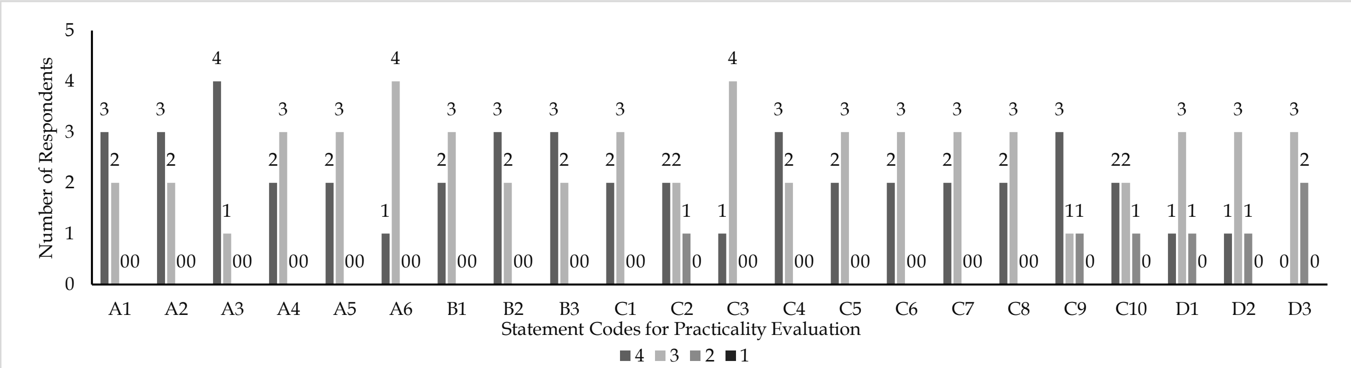


Figure 4. Graph of Practicality Evaluation Results

As for the weakness aspect, which contains negative statements, one respondent agreed with statements D2 and D2, scoring a 2 on each. From these responses, this reflects the issues mentioned in statement C2, where using a wide scoring range was considered inefficient and difficult to apply. Two respondents also agreed with statement D3, scoring a 2 for the statement. This confirms (Popham, 2017), view that having too many instrument assessment criteria can reduce effectiveness. A good instrument should be concise, targeted, and straightforward.

The practicality scores from respondents were then calculated using a formula adapted from Widoyoko (2012), by averaging all the responses. The final score was 3.334, which indicates that the perfomace-based instrument is highly practical. This shows that the performance-based assessment instrument for assessing students' SPS in the saponification lab activities is very practical and suitable for use.

Conclusion

Based on the research findings, a performance-based assessment instrument was successfully developed to assess students' science process skills (SPS). The developed instrument includes 27 performance aspects aligned with 6 SPS indicators, they are predicting, planning, observing, implementing, communicating, and interpreting. The instrument achieved a high validity score with CVI of 0.99. It also shows variable agreement results, with 2 aspects showed very strong agreement, 3 aspects showed strong agreement, and 11 aspects showed weak agreement. The practicality of the instrument resulted in a score of 3.3.5 of 4.00, indicating it is very practical. These findings suggest the need to refine several instruments items to enhance inter-rater reliability.

Acknowledgments

We would like to express our sincere gratitude to the chemistry teacher for the assistance in the data collection process during the limited trials of the instrument and for the role as a validator.

Author Contributions

The contribution of each author in this research is planning stages, Aztiannisa Zahra, Nahadi, Wiwi Siswaningsih; Methodology determination, Aztiannisa Zahra, Nahadi, Wiwi; Developing stages, Aztiannisa Zahra; Content Validity Test, Aztiannisa Zahra, Nahadi, Wiwi Siswaningsih; Pilot trial, Aztiannisa Zahra; Writing – original drafting, Aztiannisa Zahra; Writing - Reviewing and editing, Nahadi, Wiwi Siswaningsih; Supervision, Nahadi, Wiwi Siswaningsih. All authors have read and agree to the published version of the manuscript.

Funding

This research was independently funded by researchers

Conflicts of Interest

The authors declare no conflict of interest.

References

Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, *33*(9), 1289–1312. https://doi.org/10.1080/09500693.2010.512369

Ainillana, Q., & Louise, I. S. Y. (2024). Authentic Assessment Instrument on Redox Reactions to Assess Students' Cognitive Skills. *Jurnal Penelitian Pendidikan IPA*, *10*(10), 7437–7446. https://doi.org/10.29303/jppipa.v10i10.8791

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, *15*(2), 214–221.

https://doi.org/10.1016/j.sapharm.2018.03.066

Astuti, W. P. (2012). Pengembangan instrumen asesmen autentik berbasis literasi sains pada materi sistem ekskresi. *Lembaran Ilmu Kependidikan*, *41*(1), 39–43. Retrieved from http://journal.unnes.ac.id/nju/index.php/LIK

Chong, J., Mokshein, S. E., & Mustapha, R. (2021). A Content Validity Study for Vocational Teachers' Assessment Literacy Instrument (VoTAL. *International Journal of Academic Research in Business and Social Sciences*, *11*(4). https://doi.org/10.6007/IJARBSS/v11-i4/9547

Chowdhury, F. (2018). Application of Rubrics in the Classroom: A Vital Tool for Improvement in Assessment, Feedback and Learning. *International Education Studies*, *12*(1), 61. https://doi.org/10.5539/ies.v12n1p61

Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research*, *5*(4), 194–197. https://doi.org/10.1016/S0897-1897(05)80008-4

Elfrida, E., Nursamsu, N., & Ariska, R. N. (2021). Development of Performance Assessment Instruments Through Practical Learning to Improve Science Process Skills. *Jurnal Penelitian Pendidikan IPA*, 7. https://doi.org/10.29303/jppipa.v7iSpecialIssue.867

Fiolida, I. A. S., Rohaeti, E., & Suwardi. (2021). Research-Oriented Collaborative Inquiry Learning Model to Improve Students' Science Process Skills in Reaction Rate Topic. *Proceedings of the 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020*. https://doi.org/10.2991/assehr.k.210305.029

Firman, H. (2013). *Evaluasi Pembelajaran Kimia*. Bandung: Universitas Pendidikan Indonesia.

Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater Agreement and Interrater Reliability: Key Concepts, Approaches, and Applications. *Research in Social and Administrative Pharmacy*, *9*(3), 330–338. https://doi.org/10.1016/j.sapharm.2012.04.004

Hajghasem, M., Abtahi, A. R., Khalili-Damghani, K., & Yousefi-Zenouz, R. (2022). A Definition of the Level of Automation Considering ERP Systems and New Technologies to Confront the Spread of COVID-19 in Medical Centers. *Journal of Clinical Research in Paramedical Sciences*, *11*(1). https://doi.org/10.5812/jcrps-128005

Hasanah, N., Verliyanti, V., & Rokhimawan, M. A. (2020). Profesionalisme Guru Menanamkan Keterampilan Proses Sains dalam Materi IPA pada Siswa Kelas V MI Ma'arif Bego. *AULADUNA: Jurnal Pendidikan Dasar Islam*, *7*(1), 1. https://doi.org/10.24252/auladuna.v7i1a1.2020

Hasibuan, R., Adventi, F., & Rtg, R. P. (2019). Pengaruh Suhu Reaksi, Kecepatan Pengadukan, dan Waktu Reaksi pada Pembuatan Sabun Padat dari Minyak Kelapa (Cocos nucifera L. *Jurnal Teknik Kimia USU*, *8*(1). https://doi.org/10.32734/jtk.v8i1.1601

Heydarnejad, T., Tagavipour, F., Patra, I., & Farid Khafaga, A. (2022). The impacts of performance-based assessment on reading comprehension achievement, academic motivation, foreign language anxiety, and students' self-efficacy. *Language Testing in Asia*, *12*(1), 51. https://doi.org/10.1186/s40468-022-00202-4

Hikmah, N., Yamtinah, S., Ashadi, & Indriyanti, N. Y. (2018). Chemistry teachers' understanding of science process skills in relation of science process skills assessment in chemistry learning. *Journal of Physics: Conference Series*, *1022*(1). https://doi.org/10.1088/1742-6596/1022/1/012038

Irwanto, R., E., & Prodjosantoso, A. K. (2018). Undergraduate Students' Science Process Skills in Terms of Some Variables: A Perspective from Indonesi. *Journal of Baltic Science Education*, *17*(5). Retrieved from https://shorturl.asia/yGFj2

Isnaini, A. I., & Utami, L. (2020). Pengembangan Instrumen Penilaian Kinerja untuk Mengukur Kemampuan Psikomotorik Siswa dalam Praktikum Laju Reaksi. *Journal of The Indonesian Society of Integrated Chemistry*, *12*(1), 24–30. https://doi.org/10.22437/jisic.v12i1.9054

Komisia, F., Leba, M. A., & Tukan, M. B. (2022). Pelatihan Praktikum Kimia Berbasis Lingkungan untuk Meningkatkan Keterampilan Proses Sains Siswa Kelas XI MIPA SMA Negeri 12 Kupang. *Adimas Galuh*, *4*(1), 453–462. https://doi.org/10.25157/ag.v4i1.7189

Kriswantoro, K., Kartowagiran, B., & Rohaeti, E. (2021). A critical thinking assessment model integrated with science process skills on chemistry for senior high school. *European Journal of Educational Research*, *10*(1), 285–298. https://doi.org/10.12973/eu-jer.10.1.285

Kusuma, M. B., Nahadi, & Anwar, B. (2025). Preliminary Analysis for Developing Digital Assessment of Science Process Skills in Occupational Health and Safety for University Chemistry Laboratory. *Jurnal Penelitian Pendidikan IPA*, *11*(6), 7–15. https://doi.org/10.29303/jppipa.v11i6.11430

Kusumaningtyas, P., Eka, R., Dan, Y., & Majid, A. (2018). Pengembangan Instrumen Penilaian Kinerja untuk Mengukur Kompetensi Siswa dalam Kegiatan Praktikum Kimia di SMA/K. *Jurnal Inovasi Pendidikan Kimia*, *12*. Retrieved from https://journal.unnes.ac.id/nju/index.php/JIPK

/article/view/15470

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4). Retrieved from https://parsmodir.com/wp-content/uploads/2015/03/lawshe.pdf

Lestari, T., Rosidin, U., Distrik, I. W., Ertikanto, C., & Herlina, K. (2025). Analysis of Assessment Instrument Needs to Measure Students Critical Thinking Skills and Self-Efficacy in Physics Learning. *Jurnal Penelitian Pendidikan IPA*, *10*(12), 11257–11262. https://doi.org/10.29303/jppipa.v10i12.9453

Livingston, S. A. (2018). *Test Reliability-Basic Concepts (Research Memorandum No. RM-18-01*. New Jersey: Educational Testing Service.

Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and Assessment in Teaching* (10th ed.). New Jersey: Pearson Education.

Mudhakiyah, Z., Wijayati, N., Haryani, S., & Nurhayati, D. S. (2022). Pengembangan Instrumen Penilaian Aspek Psikomotorik Peserta Didik pada Praktikum Pembelajaran Kimia Materi Laju Reaksi. *Chemistry in Education*, *11*(2). Retrieved from http://journal.unnes.ac.id/sju/index.php/chemined

Murniati, S., Sudirman, & Aji, D. P. (2023). Development of Computer-Based Assessment Instruments Nuclear Physics Introduction Course as An Effort to Grade Authentically. *Jurnal Penelitian Pendidikan IPA*, *9*(7), 5103–5110. https://doi.org/10.29303/jppipa.v9i7.4416

Nahadi, & Firman. (2019). *Asesmen Pembelajaran Kimia*. Bandung: UPI Press.

Okam, C. C., & Zakari, I. I. (2017). Impact of Laboratory-Based Teaching Strategy on Students' Attitudes and Mastery of Chemistry in Katsina Metropolis. *International Journal of Innovative Research and Development*, *6*(1). https://doi.org/10.70771/jocw.v2i2.30

Oktavilani, N. D., & Agustini, R. (2024). Development of Performance Assessment Instrument to Measure Students Laboratory Skills in Thermochemistry Practicum Exothermic and Endothermic Reactions. *Jurnal Pijar Mipa*, *19*(5), 786–791. https://doi.org/10.29303/jpm.v19i5.7069

Popham, W. J. (2017). *Classroom assessment : what teachers need to know* (8th ed.). Los Angeles: Pearson.

Putri, D., Supriatna, A., & Rahmawati, T. (2024). Soap making project from waste cooking oil for high school students' chemistry learning: qualitative content analysis. *Jurnal Penelitian Pendidikan IPA*, *10*(6), 3147–3154. https://doi.org/10.29303/jppipa.v10i6.7701

Rahayu, A. (2020). Analysis of Students' Science Process Skills in Practicum on the Basics of Analytical Chemistry. *Jurnal Pendidikan Kimia Dan Ilmu Kimia*, *3*(1). https://doi.org/10.31602/dl.v3i1.3102

Rasyidin, & Mansur. (2009). *Penilaian Hasil Belajar*. Bandung: CV Wacana Prima.

Rosidin, U., Herliani, D., & Viyanti. (2023). Development of Assessment Instruments in Project-Based Learning to Measure Students Scientific Literacy and Creative Thinking Skills on Work and Energy Materials. *Jurnal Penelitian Pendidikan IPA*, *9*(6), 4484–4494. https://doi.org/10.29303/jppipa.v9i6.2421

Saragih, D., Silaban, R., & Darmana, A. (2021). Pengembangan modul makromolekul berbasis proyek dengan pendekatan sets (science, environment, technology and society. In *Prosiding Seminar Nasional Kimia* (pp. 35–41). Retrieved from https://jurnal.kimia.fmipa.unmul.ac.id/index.php/prosiding/article/view/1039

Sari, R. P., Mauliza, M., Nazar, M., & Nahadi, N. (2020). The Implementation of Performance Assessment Through Virtual Laboratory to College Students' Creative Thinking Skills. *Jurnal Penelitian Pendidikan IPA*, *7*(1), 5. https://doi.org/10.29303/jppipa.v7i1.484

Schellekens, L. H., Bok, H. G. J., Jong, L. H., Schaaf, M. F., Kremer, W. D. J., & Vleuten, C. P. M. (2021). A scoping review on the notions of Assessment as Learning (AaL) Assessment for Learning (AfL), and Assessment of Learning (AoL). *Studies in Educational Evaluation*, *71*, 101094. https://doi.org/10.1016/j.stueduc.2021.101094

Sejati, F., Saputro, S., & Indriyanti, N. (2020). Analysis of Students' Science Process Skills on the Concept of Material Classification and its Changes in the Junior High School. *International Journal of Educational Research Review*, *5*(1). https://doi.org/10.24331/ijere.666587

Setiono, & Astuti, F. O. (2021). Mengajarkan Keterampilan Proses Sains dalam Pembelajaran Jarak Jauh. *BIO-EDU: Jurnal Pendidikan Biologi*, *6*(2), 74–82. https://doi.org/10.32938/jbe.v6i2.1121

Sibic, O., & Şesen, B. A. (2022). Examining Science Process Skills Tests: A Case of Turkey. *International Journal of Assessment Tools in Education*, *9*(1), 20–38. https://doi.org/10.21449/ijate.1058055

Siswaningsih, W., Kusrijadi, A., Pitriyani, R., & Pratiwi, M. M. (2024). Performance Assessment Model in Experimental Electrolysis for High School Student. *Jurnal Pengajaran Matematika Dan Ilmu Pengetahuan Alam*, *28*(2), 48–53. https://doi.org/10.18269/jpmipa.v28i2.65005

Sofyan, A., Feronika, T., & Milama. (2019). *Evaluasi Pembelajaran Sains berbasis Kurtilas*. Jakarta: Yasmin.

Stephenson, N. S., Duffy, E. M., Day, E. L., Padilla, K., Herrington, D. G., Cooper, M. M., & Carmel, J. H. (2020). Development And Validation Of Scientific Practices Assessment Tasks For The General Chemistry Laboratory. *Journal Of Chemical Education*, *97*(4), 884–893. https://doi.org/10.1021/Acs.Jchemed.9b00897

Sudirman, S., Ramdani, A., Doyan, A., Anwar, Y. A. S., Rokhmat, J., & Sukarso, A. A. (2023). Case Study in West Nusa Tenggara for Automated Feedback of Performance Assessment on Science Practicum to Measure Science Process Skills in University. *Jurnal Penelitian Pendidikan IPA*, *9*(12), 11903–11910. https://doi.org/10.29303/jppipa.v9i12.6370

Sugrah, N., Abu, S. H. N., Rahman, N. A., Rakhman, K. A., Danial, M., & Anwar, M. (2019). Assessment of Processes and Resources for Knowledge of Skills of a Chemistry Laboratory at the Senior High School of Ternate Island. *Journal of Curriculum and Teaching*, *8*(1), 11–19. https://doi.org/10.5430/jct.v8n1p11

Sullivan, G. M. (2011). A Primer on the Validity of Assessment Instruments. *Journal of Graduate Medical Education*, *3*(2), 119–120. https://doi.org/10.4300/jgme-d-11-00075.1

Viyanti, R., L., D., & Rosidin, U. (2023). Development of Assessment Instruments in Project-Based Learning to Measure Students Scientific and Numeracy Literacy Ability on Harmonic Vibration Materials. *Jurnal Penelitian Pendidikan IPA*, *9*(7), 5039–5048. https://doi.org/10.29303/jppipa.v9i7.2408

Widoyoko, E. P. (2012). *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Pelajar.

Wolf, K., & Stevens, E. (2007). The role of rubrics in advancing and assessing student learning. *Journal of Effective Teaching*, *7*(1), 3–14. Retrieved from https://files.eric.ed.gov/fulltext/EJ1055646.pdf

Wulan, R. A. (2018). *Menggunakan Asesemen Kinerja untuk Pembelajaran Sains dan Penelitian*. Bandung: UPI Press.

Yusup, F. (2018). Uji validitas dan reliabilitas instrumen penelitian kuantitatif. *Tarbiyah: Jurnal Ilmiah Kependidikan*, *7*(1). https://doi.org/10.18592/tarbiyah.v7i1.2100