Available online at: http://journal.uny.ac.id/index.php/jipi

Jurnal Inovasi Pendidikan IPA, 10 (1), 2024, 32-46

Developing and Validating a Rubric for Measuring Skills in Designing Science Experiments for Prospective Science Teachers

Rizki Nor Amelia*, Prasetyo Listiaji, Novi Ratna Dewi, Andhina Putri Heriyanti, Bagus Dwi Atmaja, Tafuz Mahabatis Shoba, Imam Sajidi

Universitas Negeri Semarang. Semarang, Indonesia

*Corresponding Author. E-mail: rizkinoramelia@mail.unnes.ac.id

Received: 09 September 2023; Revised: 11 Januari 2024; Accepted: 25 April 2024

Abstract: The Indonesian government mandates that science teachers must have competence in designing science experiments for learning purposes so that science content can be learned optimally by students while preparing them to have the ability to face the 21st century. This is development research that aims to develop a measurement instrument for science experiment design skills for prospective science teachers that meets good psychometric characteristics. The rubric development procedure refers to the Churches rubric development method, which consists of four stages: define, design, do, and debrief, involving 10 experts (lecturers and teachers) and 124 prospective science teachers as research participants. The results of exploratory and confirmatory factor analysis showed that the analytical rubric developed by measuring ten aspects, namely title, research objectives, relevant theories, variables, materials, equipment and instrumentation, method, an appropriate number of data, references, and systematic and technical writing was valid in content (CVI=.96), valid in construct (GFI=.94; RMSEA=.071; NFI=.99; CFI=1.00; PNFI=.91), and reliable (α=.968). The use of a standardized rubric certainly allows the assessment to provide consistent, accurate, and objective results and helps students understand what competencies they must achieve.

Keywords: prospective science teacher, rubric, science experiment design skills

How to Cite: Amelia, R.N., & Listiaji, P., Dewi, N.R., Heriyanti, A.P., Atmaja, B.D., Shoba, T.M., & Sajidi, I. (2024). Developing and Validating a Rubric for Measuring Skills in Designing Science Experiments for Prospective Science Teachers. *Jurnal Inovasi Pendidikan IPA, 10*(1), 32-46. doi: http://doi.org/10.21831/jipi.v10i1.65853

INTRODUCTION

Every teacher, including science teachers, has to provide meaningful learning experiences to students that aim to develop the skills needed to face various challenges in life. In addition, the rapid development of technology and the digital era in the 21st century will affect globalization and increase competition among people. Preparing humans to compete in the 21st century can be done in various ways, one of which is through education, so strengthening character in schools aims to develop student character following the core competencies needed in the 21st century, namely critical thinking and problem-solving skills, communication skills, collaboration skills, and creativity and innovation (Petrie, 2023). To achieve this, science learning demands mastery of science process skills to advance the level of thinking and support learning skills needed to face the 21st century (Putri et al., 2022), especially those related to investigative and exploratory activities as scientists in general (Wola et al., 2023).

Science process skills are defined as a person's skills to use thought, reason, and effective action effectively and efficiently (C. A. Dewi et al., 2019) so that they can stimulate activeness, facilitate scientific learning, develop responsibility, enhance the durability of learning, and provide research methodologies (Gürses et al., 2015 & Kurniawati, 2021) which is indispensable in the development of science and daily life (Indrawati, 2017). These skills are divided into two major: basic process skills (i.e., observing, inferring, measuring, communicating, classifying, predicting, and using integrated space-time and numbers) and integrated process skills (i.e., controlling variables, defining operations, defining data, and analyzing data). controlling variables, defining operationally, formulating hypotheses, formulating models, interpreting data, and experimenting) (Kurniawati, 2021), which, when trained with laboratory-based experimental learning can serve as a foundation for other cognitive skills such as logical thinking, reasoning, and problem-solving (Kanapeckas Métris, 2020). Since long ago, there have been many studies that explain the importance of the role of laboratories in science education (Hofstein & Mamlok-Naaman, 2007), one of which is the design of scientific notations and models for testing hypotheses (Shana & Abulibdeh, 2020).

Experiments are needed to deliver science materials because they make students understand and grasp science concepts better (Malik, 2018), improve understanding and develop skills in solving problems, and understand the nature of science by replicating the actions of scientists (Shana & Abulibdeh, 2020). Science experiments can be done in a real laboratory (using KIT or tools available in everyday life) or a virtual laboratory (i.e., PHET simulation, chem collective, labster) (Malik, 2018). Practicing process skills in their implementation certainly begins with teacher modeling, and then students are asked to work and practice according to the teacher's instructions and guidance (Heryani et al., 2023). Teachers play an important role in facilitating science process skills in the classroom through planning and organizing teaching and learning activities to achieve scientific information (C. A. Dewi et al., 2019). Therefore, prospective science teachers need to have skills in designing effective and efficient science experiments (Pahrudin et al., 2019).

One form of measurement instrument that can be used to measure prospective science teachers' skills in designing science experiments is a rubric (Rukmini & Saputri, 2017). A rubric is an assessment tool that has a description of the expected performance of each criterion to achieve a certain value or result (Cooper, 2023). Some experts explain the benefits of rubrics, including saving time and speeding up the feedback process (Hettithanthri et al., 2023), improving performance to achieve a set of standards (Sadler, 2009), enabling supervision and monitoring of student progress (Reddy & Andrade, 2010), helping students to focus on their learning efforts, producing higher quality work and assignments to achieve better grades (Reddy & Andrade, 2010), and providing more accurate, fair, and transparent assessments that can avoid personal prejudice (Isbell & Goomas, 2014). The two most frequently encountered rubrics are the analytic rubric and the holistic rubric (Sadler, 2009).

In analytic rubrics, each dimension or criterion is evaluated separately, whereas in holistic rubrics, all dimensions are evaluated simultaneously (Cooper, 2023). Although analytic rubrics are good for formative assessment, it takes more time to grade the assignment compared to holistic rubrics. In contrast, assessment can be faster with holistic rubrics, and this type of rubric is also suitable for summative assessment. However, holistic rubrics provide a single overall score that fails to provide clear information on where or how improvement can be achieved by the user (Chowdhury, 2018). The choice between the two types of rubrics can be customized to accommodate the preferences of evaluators, the nature of the assignments provided, or the educational goals and competencies (Sudaryanto & Akbariski, 2021).

Some research related to science experiments designed by teachers or prospective teachers has been done, even though the number is still very limited. Research conducted by (C. A. Dewi et al., 2019) in the form of a comparative study of differences in the effect of understanding concepts and mastery of science process skills on the ability to design science experiments found that understanding science concepts (cognitive ability, insight, or knowledge of science) and mastery of basic science process skills (observing, classifying, measuring, interpreting data, predicting, experimenting, and concluding) both have a positive and significant effect on Roudhotul Athfal and Madrasah Ibtiyah prospective teachers. The limitation of this study is documentation techniques (photos and videos of experiments) for data collection, complemented by Likert scale assessment instruments to assess concept understanding and mastery of process skills, and conducting interviews to strengthen and support research data.

Research by Davy Tsz Kit et al (2022) to improve the competence of the teachers of senior high school (SMA) and madrasyah aliyah (MA) in designing physics experiments as an effort to train 21st-century skills is done by providing insight into how to design experiments, types of experimental designs (i.e., inquiry experiments, problem-solving, conceptual change), and virtual experiment training. The findings of this research show that the training activities on designing experiments can provide teachers with knowledge and motivation to carry out experiments in physics learning at school so that it is

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

expected to train students' 21st-century skills. In this study, the focus was limited to providing training, and then the training was evaluated using an open-ended questionnaire to gather information about participants' responses to the training activities and a 10-item multiple-choice test to gather information about concept mastery in designing experiments.

Similar research was conducted by Pursitasari et al (2023), who provided training on the preparation of experiment designs and the use of science experiment kits for science teachers of junior high schools in Serang Regency. In this study, the design of science experiments made by teachers was assessed with a rubric (scores 1-4) with assessment components namely title, objectives, literature review, tools and materials used, how to work, and references. However, the assessment instrument used was not clearly explained in terms of its psychometric characteristics. This study focused on discussing whether the training activities were able to improve teachers' science process skills (i.e., problem identification, planning, observation, prediction, interpretation, and writing the results of the practicum well). Even so, there are interesting findings, namely that the experiment design made by 37.5% of teachers is SDG-oriented, 42.5% is context-based, and 20% is still a cookbook. The cookbook experiment design shows the existence of detailed and clear objectives, tools and materials, and ways of working so that students only experiment according to the work steps contained in the student worksheet. Whereas the cookbook experiment design has not facilitated students' development of science process skills (Sari & Zulfadewina, 2020), it is better if students are trained in their scientific abilities gradually through a semi-open-ended experiment and then increase to an open-ended experiment.

Based on the review of previous research, there are limitations in the form of the absence of a complete description of the psychometric characteristics of the instruments used to measure the ability to design science experiments of prospective teachers. This means that no one has examined the development of instruments that function to measure skills in designing experiments, especially on the subject of prospective science teachers. The psychometric characteristics of the instrument are very important because these characteristics are the identity of the measurement instrument which will be directly proportional to the measurement data produced. If the characteristics of the instrument are not known or even not analyzed, it is not wise if the data generated from the instrument is made into a research conclusion to generalize to the population. The development of this measurement instrument certainly has the potential to analyze the strengths and weaknesses of prospective science teachers in designing experiments, where this skill is one component of the competencies that must be possessed by science teachers of junior high school as mandated in the Regulation of the Minister of National Education of the Republic of Indonesia Number 16 of 2017 concerning Standards for Academic Qualifications and Teacher Competencies, which is specifically written in the twelfth point, namely designing science experiments for learning or research purposes..

METHOD

Development Procedure

This study is development research that aims to develop a rubric that meets good psychometric characteristics. The rubric development was carried out by following (Mang et al., 2023) 4Ds rubric development steps, which consist of four stages: define, design, do, and debrief. In general, the four steps of rubric development are defined as follows. The defined stage is the stage of formulating the assessment objectives that must be achieved and identifying the key elements or components that will be assessed using the rubric to be developed. At this stage, the assessment objectives must be clearly defined, and the structure that needs to be considered when formulating objectives is SMART (specific, measurable, achievable, realistic, and timely). In addition, it is necessary to confirm whether the measurement objectives to be carried out are following the curriculum referred to. Does it link to learning objectives? Is it suitable for the audience? Can students this age achieve at the highest level? And ensure that each aspect has a standard that serves to describe the skill (i.e., excellent standard, good standard, acceptable standard, poor standard, and failing).

The design stage is divided into two sub-stages, namely assessment mode and assessment design. The assessment mode sub-stage is a stage that ensures the assessment model used (i.e., summative or formative), when this instrument can be used, whether it can be used with students or only teachers can use it; while the assessment design sub-stage is where the format and structure of the rubric are

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

determined (i.e., ascending 1 to 4 or descending 4 to 1). While this phase is brief, it holds significant importance. It involves deciding on the rubric's style and the specific command terms to be used. The Do stage (rubric development) is the stage of preparing the rubric items. At this stage, the rubric of science experiment design skills was developed in 10 items (see Table 1), each of which measured one aspect and consisted of four descriptors for each aspect, where a score of 4 indicates that four aspect indicators have been met, a score of 3 indicates that one aspect indicator has not been met, a score of 2 indicates that two aspect indicators have not been met, a score of 1 indicates that only one aspect indicator has been met, and a score of 0 indicates that none of the aspect indicators has been met.

Finally, the debrief stage (use and evaluate) begins with expert validation. An expert should ensure clarity regarding the validity of an instrument, which is described as the extent to which a measuring tool accurately assesses what it is intended to measure or fulfills its intended purposes (Freeman & Jessup, 2004). Therefore, experts, recognizing the significance of their role and understanding that the validation process will be accompanied by an immediate assessment of reliability, should rigorously scrutinize each component of the instrument (Pino et al., 2023). With the help of experts, matters related to the content of key aspects, potential items relevant to the construct, item clarity, language complexity, and other item issues that may have escaped the researcher will be re-examined (Fatiyah et al., 2021).

Participants

The participants involved in this study are lecturers, teachers, and prospective science teachers. A total of seven lecturers and three teachers acted as expert validators to assess the assessment rubric developed. The educational background of the validators is as follows: one doctor, one doctor candidate, six masters, and two bachelors; who teach at five universities and three private schools; while pre-service science teacher students in semesters 4 and 6 comprise a total of 124 people, established by using the cluster random sampling technique, as a party who develops a science experiment design and presents it in the form of a student worksheet, which is then used as material to prove empirical validity. The science experiment design developed refers to science material in the Merdeka Curriculum that provides students with the flexibility to choose lessons according to their interests (Ploj Virtič, 2022).

Data Analysis

The first analysis was to prove content validity. The lecturer and teacher rated the Science Experiment Design Skill (SEDS) Rubric content with three scales, namely "essential, useful but not essential, and not necessary," related to the construct, which was then processed with content validity ratio (CVR) and content validity index (CVI) statistics using (Lawshe, 1975) formulas as equations (1) and (2), where *n* is the number of panelists indicating "essential" and N is the total number of panelists.

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}} \ldots (1) \text{ and } CVI = \frac{CVR}{N} \ldots (2).$$

After the rubric was proven to be content valid and followed by improvements according to the validator's suggestions, its empirical validity was then proven through two stages of construct validity, namely exploratory factor analysis (EFA) in the first stage and confirmatory factor analysis (CFA) in the second stage. The reliability is estimated using Cronbach's alpha formula (Creswell, 2014) and construct reliability (Cheung et al., 2023) as equations (3) and (4).

$$\alpha = \frac{K}{K-1}\left[1 - \frac{\sum S^2 y}{S^2 x}\right] \ldots (3)$$

$$CR = \frac{\left[\sum_{i=1}^{n} \lambda_i\right]^2}{\left[\sum_{i=1}^{n} \lambda_i\right]^2 + \sum(1-\lambda_i^2)} \ldots (4)$$

where: α = Cronbach Alpha, K = number of test items, $\sum S^2 y$ = sum of item variance, and $S^2 x$ = variance of total score, CR = construct reliability, and λ_i = standardized factor loading item-i

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

FINDING AND DISCUSSION

As explained earlier, the main focus of this research is to develop a SEDS rubric measurement instrument for prospective science teachers that meets good psychometric characteristics. The first psychometric characteristic examined was content validity using Lawsche's formula. Based on the formula, the CVR cutoff for 10 validators was set at .62 (Lawshe, 1975). The results of the content validity analysis on the developed measurement instruments showed that the CVR value moved between .8 and 1.00. The CVR value of 0.8 indicates that there is one validator who considers that items 8 (appropriate number of data) and 9 (references) are useful but not essential to be used as a reference for assessing science experiment design skills for prospective science teachers. Appropriate numbers of data are considered not essential with the consideration that the science experiments developed do not require a lot of data variation because the target users are junior high school students, while references are also considered not essential with the consideration that the list of references is not a major component in the systematics of the science worksheet (Cholifah & Novita, 2022). Nevertheless, the reference aspect is still included as one of the components of assessing the ability to design science experiments because references can make the teaching materials developed richer in information (Heryani et al., 2023). So, in general, it can be concluded that all items have CVR values above the cutoff set with an average CVI of .96 (see Table 1), which means that all items have proven to have good content validity.

Table 1. Final SEDS rubric

No.	Aspect	Descriptor	CVR
1.	Title	a. A short title, consisting of a maximum of two lines with a maximum word count of 10–12 words. b. The title is clear, informative, specific, and relevant to the experiment to be conducted. c. The title is not an abbreviation. d. The title is packed with keywords that are easily searchable.	1.00
2.	Research objectives	a. Objectives are appropriate to the subject matter. b. Objectives are conceptually correct. c. Objectives are formulated in a SMAR (specific, measurable, achievable, and relevant) manner. d. Objectives are written in concise and easy-to-understand sentences.	1.00
3.	Relevant theory	a. The theory referred to is conceptually correct by including a reliable reference source and citing it using appropriate techniques. b. The theory supports the experiment to be conducted. c. The theory is well written and includes equations (if any) and discussion relevant to the experiments. d. Theory, published data on similar experiments, or simulations relevant to the experiment are used to predict the results.	1.00
4.	Variables	a. Variables are operationally defined. b. Variables (independent, dependent, control) are appropriately identified. c. Variable ranges or values are identified as per theory or published data or simulations relevant to the experiment. d. The relationship between variables is clear.	1.00
5.	Materials/specimens	a. Materials are easily available. b. The type, size, amount, and/or concentration of materials used are appropriate for the needs of the experiment. c. The materials used are safe, e.g. non-flammable, non-explosive, non-irritating, non-polluting to the environment, and non-toxic.	1.00

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

No.	Aspect	Descriptor	CVR
		d. Materials are in good condition, e.g. not contaminated and not expired.	
6.	Equipment and instrumentations	a. Tools are easily available. b. Tools are easy to use (or there is a manual book). c. The type, size, and/or number of tools used are appropriate for the needs of the experiment. d. The tool is in good condition, e.g., the components are complete, sensitive, and precise to make measurements	1.00
7.	Method	a. Procedures are correct and follow science concepts. b. Procedure is written in logical steps with a numbered sequence or depicted in a flow chart. c. Procedure is effective or appropriate for the experiment. d. Procedure is efficient or does not require a long experimental time.	1.00
8.	Appropriate number of data	a. The amount of variation in the measured data follows theoretical considerations, experimental kit compatibility, availability of materials and experimental time, and potential errors. b. A clear table to record all data obtained during the experiment is provided. c. Data analysis methods are appropriate for the experiment. d. Data analysis steps are presented in a structured and easy-to-understand manner.	.80
9.	References	a. All references cited are listed in the reference list. b. References must be from reliable secondary sources. c. References are current (last 3-5 years). d. References are written using a specific format (APA, Vancouver, MLA, etc.).	.80
10.	Systematic and technical writing	a. The systematic writing of student worksheets is organized according to the format, starting from the title to the reference list. b. Student worksheet is equipped with an evaluation sheet containing questions that are relevant to experimental activities. c. Every sentence is effective and uses the Indonesian language that follows PUEBI. d. There are no spelling errors or typos, and the appropriate type, font size, and margins are used.	1.00
		CVI	.96

After all aspects have been confirmed to be content valid, the next step is to prove construct validity because there are no latent variables previously constructed, so it is necessary to conduct EFA to identify the number of latent variables and continue with confirmation through CFA (Dash & Paul, 2021). Construct validity provides information or data to prove that the items in the scale are correlated and together measure the construct they are meant to measure (Svenningsson et al., 2022), which can be proven through EFA and CFA. EFA is employed to unveil the underlying structures of variables that encompass various components, the complete configuration of which may not be entirely understood but is recognized to exist and essential for each item within a measurement instrument to assess a characteristic for ensuring validity (Acar Güvendir & Özer Özkan, 2022), whereas CFA is used to validate or refute a theory that pertains directly to what the instrument is gauging (Cheung et al., 2023). In simple terms, both outcomes of the analysis would provide proof regarding the suitability of the items within the SEDS rubric within the theoretical framework.

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

We began the EFA by examining a total of 10 items that measure 10 different aspects of SEDS. These aspects include title, research objectives, relevant theories, variables, materials and specimens, equipment and instrumentation, method, appropriate number of data, references, systematics, and technical writing. In EFA, three things first need to be considered to measure the adequacy of sampling before data can be properly factored: the sample size estimated by the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO-MSA), the identity matrix with Bartlett's Test of Sphericity, and the factorability of the correlation matrix (Timm & Barth, 2021). KMO values < .6 indicate inadequate sampling and remedial action should be taken, while KMO-MSA < .5 indicates that the results of the factor analysis undoubtedly will not be very suitable for the analysis of the data (Fields et al., 2021). In addition to describing general sampling adequacy through KMO-MSA, sampling adequacy can also be evaluated on individual items by referring to anti-image correlation statistics >.5 (Wu et al., 2023).

Bartlett's Test of Sphericity evaluates whether the variables exhibit orthogonality, meaning that the initial correlation matrix resembles an identity matrix, implying no connection between the variables and rendering them inadequate for structural analysis (H_0). Conversely, the alternative hypothesis (H_a) suggests that the variables are not orthogonal, indicating substantial divergence between the correlation matrix and the identity matrix due to correlations among them. A significance level of less than .05 signifies the potential suitability of conducting a factor analysis on the dataset (Rodriguez et al., 2019). Based on this reference, the statistical requirements for this study were met successfully (see Table 2), so the factor analysis can be continued by checking the factorability of the correlation matrix in the component matrix (see Table 3).

Table 2. KMO Dan Bartlett's test

		SEDS Rubric
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.891
Bartlett's Test of Sphericity	Approx. Chi-Square	1241.650
	df	45
	sig.	.000

The results of factor analysis in the component matrix show that the SEDS rubric instrument developed is proven to measure only one factor (unidimensional), namely science experiment design skills. This is evidenced by the high loading factor of each aspect (.810-.951) with a satisfactory total variance explained value of 78.468%; it is recommended that the retained factors account for a minimum of 50% of the total explained variance (Rodriguez et al., 2019). Evidence of unidimensionality can also be interpreted from the scree plot (Ledesma et al., 2015) presented in Figure 1. Scree plot is a heuristic graphical approach involving two steps: first, plotting eigenvalues (on the y-axis) against components (on the x-axis); and second, examining the shape of the resulting curve to identify the point where a significant change occurs (Ledesma et al., 2015). Factor extraction should be stopped at the juncture where a distinct "bend" or plateau appears in the plot. This assessment is employed to determine the ideal number of factors that can be extracted before the influence of unique variance surpasses the shared variance structure (Cooper, 2023). The results of the scree plot (see Figure 1) show that there is one point at the bend of the elbow that is significant, so the measurements made with the SEDS rubric are unidimensional.

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

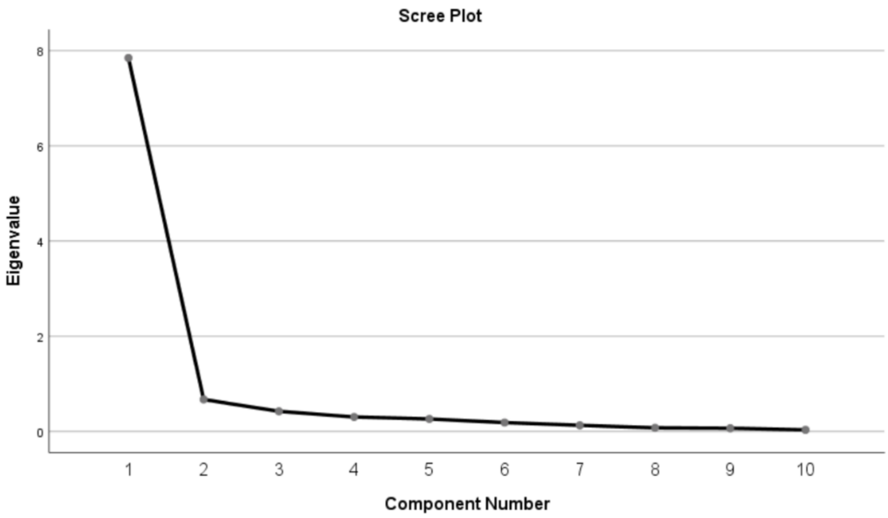

Figure 1. Scree plot SEDS rubric

Meanwhile, evidence of the reliability of the instrument was estimated using Cronbach's alpha formula, where the alpha value ranged between 0 and 1, with 7 considered the minimum acceptable value (Viñas, 2022). Table 3 shows that the rubric for assessing science experiment design skills proved to be reliable, with an alpha reliability coefficient of .968. In the same table, the corrected item-total correlation value is also presented, which serves as a measure of the consistency of the adjusted items, where if the value is equal to zero or even negative, it should be considered for deletion, and the Cronbach's alpha if item deleted value, which reflects the alpha value when certain items are removed (Viñas, 2022). These two references allow instrument developers to confirm that eliminating items with the weakest correlation to the total score does not result in a reliability coefficient that is significantly higher than the overall reliability. In this study, no items needed to be removed because all items had corrected item-total correlation values that met the criteria.

Table 3. Exploratory factor analysis and alpha reliability estimation of SEDS rubric

Items	Anti Image Correlation	Component Matrix	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Title	.871	.847	.804	.966
Research objectives	.868	.930	.914	.962
Relevant theory	.911	.828	.787	.966
Variables	.949	.930	.908	.962
Materials/specimens	.842	.922	.898	.962
Equipment and instrumentations	.873	.946	.926	.962
Method	.922	.951	.938	.961
Appropriate number of data	.919	.844	.809	.966
References	.927	.833	.796	.967
Systematic and technical writing	.846	.810	.771	.967
Total Variance Explained	78.468%		**Cronbach Alpha**	.968

The construct description of the SEDS measurement model generated through EFA was further verified for factorial validity with CFA. This is because CFA can provide further evidence of the suitability of the suggested model concerning the factor structure identified through EFA (Hidayat et al., 2018), based on three important results, namely parameter estimation, fit index, and modification index (Gebremedhin et al., 2022). In parameter estimation, a higher factor loading makes the variable more representative of the factor (Setchell, 2019). The variable suitability for inclusion in the model should generally exhibit factor loadings greater than 0.60, although in certain cases, it might be

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

justifiable for them to be lower (Afthanorhan et al., 2020). Based on these criteria, it can be concluded that all estimated parameters have appropriate and significant λ values (indicated by blue lines, not red lines; see Figure 2 and Table 4). This means that all parameters tested are indicators of SEDS latent variables that have proven significant at the 5% significance level. Although it has met the criteria related to factor loadings, the fact is that the SEDS measurement model tends not to fit in terms of model fit statistics (see Table 4), meaning that the model created has not been able to represent well the relationship found in the sample, or it can be said that the model is not consistent with the relationship that occurs in the actual data. Therefore, modifications need to be made to obtain better test results and meet the general rules suggested for the feasibility of a measurement model.

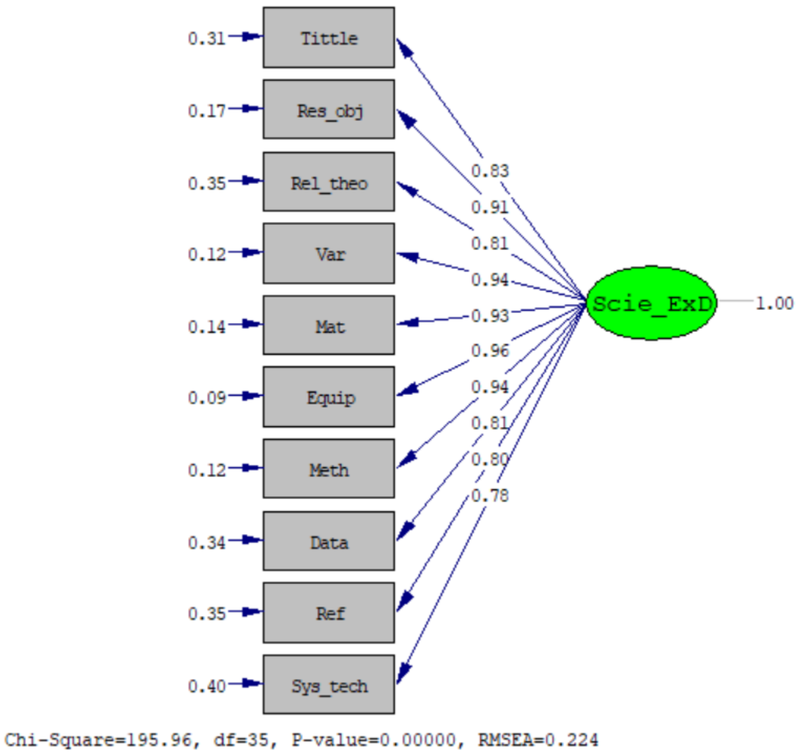

Figure 2. The original construct of the SEDS rubric

In cases where there is a λ that has a value below the cutoff, the modification can be done by removing the estimated parameter that has a λ value below the cutoff (Geldhof et al., 2014). However, the results of testing the SEDS construct in this study are not the case, so the modification is done by following the advice of the Modification Indices (MI) (Kurukunda et al., 2020). Modification indices are defined as χ^2 statistics with df = 1 (Álvarez et al., 2013), where each MI value conservatively estimates the extent to which χ^2 is reduced if a particular parameter is included in the model. Modifications can be made when the decrease in the χ^2 value is at least 3.84, as this is the critical χ^2 value at df = 1, and by changing fixed parameters to estimated (free) parameters to achieve maximum improvement in model fit (Erduran et al., 2021). LISREL reports MI, both in the output and on the path diagram (by selecting MI from the estimation menu), in the form of numbers that offer suggestions for improving overall model fit (Lee et al., 2013). Based on LISREL's suggestions, 16 MIs can be used as a basis for modifying the model by giving the Set Error Covariance Free command to the two parameters to be connected.

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

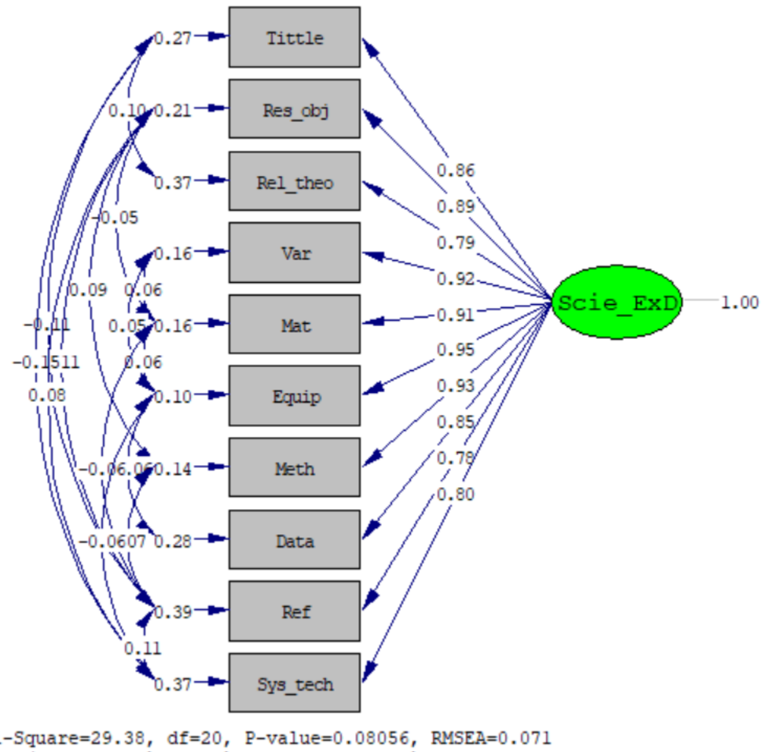

Figure 3. Final construct of SEDS rubric

Table 4. Factor loading (λ), measurement error (δ), and t-value of original-final SEDS construct

Parameter	Original construct			Final construct		
λ	**λ**	**δ**	**t-value**	**λ**	**δ**	**t-value**
λ_1	.83	.31	9.83*	.86	.27	10.21*
λ_2	.91	.17	11.34*	.89	.21	10.91*
λ_3	.81	.35	9.36*	.79	.37	9.13*
λ_4	.94	.12	11.94*	.92	.16	11.41*
λ_5	.93	.14	11.73*	.91	.16	11.39*
λ_6	.96	.09	12.41*	.95	.10	12.14*
λ_7	.94	.12	12.04*	.93	.14	11.75*
λ_8	.81	.34	9.5*	.85	.28	10.07*
λ_9	.80	.35	9.33*	.78	.39	8.91*
λ_{10}	.78	.40	8.84*	.80	.37	9.10*

Note: *p<.05

The overall model fit can be assessed by different model fit indices. In this study, the model fit indices referred to are absolute fit (i.e., GFI, RMSEA), incremental fit (i.e., CFI, NFI), and parsimonious fit (i.e., PNFI). Absolute fit indices are metrics that originate from the fit of observed covariance matrices and do not require an alternative model as a reference point for comparison (Dash & Paul, 2021). These indices reveal which of the proposed models best aligns with all available models and ascertain the compatibility between sample data and a pre-established model (McDermott-Dalton, 2022). On the other hand, incremental fit indices gauge the incremental, comparative, or relative fitness of a model when compared to a null model (Dash & Paul, 2021); and parsimonious fit indices, which belong to the relative fit indices category, are derived from the two aforementioned types and penalize complex models and encourage simpler ones (Cooper, 2023).

Absolute fit is assessed through the Goodness of Fit Index (GFI), which offers an alternative to the χ^2 test by estimating the proportion of the variance explained by the projected covariance within the population (Dash & Paul, 2021), and the RMSEA, which assesses how far a hypothesized model is

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

from a perfect model (Xia & Yang, 2019). Incremental fit is assessed through the Normed Fit Index (NFI), which evaluates the model by contrasting the χ^2 value of the model with the null model (the null mode represents the least favorable scenario, suggesting that all measured variables are unrelated) (Han et al., 2022), and the Comparative Fit Index (CFI), which is a revised NFI specifically designed to accommodate small sample sizes (Dash & Paul, 2021). Lastly, the parsimonious fit is assessed through the Parsimonious Normed Fit Index (PNFI), which is derived from the NFI by adjusting for the loss of degrees of freedom (Setchell, 2019). The modification results showed an increase in all model fit statistics and a decrease in the value of the χ^2 statistic (χ^2/df(35) original = 195.96 to χ^2/df(20) final = 29.38), so it can finally be concluded that the SEDS measurement model proved to be a good fit (see Figure 3 and Table 5).

Table 5. Goodness of fit test of SEDS rubric

Goodness of Fit Index	Cut off value	Original construct		Final construct	
		Scores	Fitness	Scores	Fitness
Absolute Fit					
GFI	GFI ≥ .90	.70	Misfit	.94	Fit
RMSEA	RMSEA ≤ .08	.22	Misfit	.071	Fit
Incremental Fit					
NFI	NFI ≥ .90	.91	Fit	.99	Fit
CFI	CFI ≥ .95	.93	Fit	1.00	Fit
Parsimonious Fit					
PNFI	PNFI ≥ .50	.45	Misfit	.91	Fit

The exploration of psychometric characteristics after the fit of the measurement model through CFA is the estimation of construct reliability (CR), which is a measure of the internal consistency of the variables representing the latent constructs that have been measured (Chen et al., 2023). Construct reliability can indeed be estimated after construct validity is proven using confirmatory factor analysis on a suitable measurement model (Geldhof et al., 2014), but CR is not appropriate if CFA produces a multidimensional measurement model because they ignore the second-order factor structure (Svenningsson et al., 2022). A construct is declared to have good construct reliability when the CR value is >.7 (Rusilowati et al., 2018); thus, the cumulative error variance should be below 30% of the variance in the latent variable (Cheung et al., 2023) because low reliability in the underlying scales increases the standard errors of estimated parameters, resulting in less powerful testing (Erduran et al., 2021). In this study, the reliability estimation results were very satisfactory, with a value of 0.969.

In the end, the goal of this study, which is to develop a SEDS rubric measurement instrument for prospective science teachers that meets good psychometric characteristics, was successfully achieved, so that it can be utilized for research purposes and related institutions. Teacher competence in science experimentation is certainly one of the main factors. With good competence, teachers will be able to optimally utilize available facilities, identify and design appropriate experimental activities, and create simple experimental tools that can be utilized in the teaching process, which ultimately contributes to improving students' science process skills (Purnamasari, 2020), students' critical thinking abilities (Bahtiar et al., 2022), students' scientific attitudes (Citra Ayu Dewi et al., 2021), and students' science literacy (Jayanti & Nurfathurrahmah, 2023).

CONCLUSION

Designing science experiments for teaching purposes is one of the core competencies that science teachers must have, as mandated by the Indonesian government. This research has successfully verified a rubric to measure prospective science teachers' skills in designing science experiments. Using a standardized rubric certainly allows the assessment to provide accurate and objective results. Standardized rubrics can be used to effectively measure the extent to which researchers' interventions have made a difference. It is important to develop educational programs to promote science experiment design skills so that teachers can equip their students to advance their level of thinking and support the learning skills needed for the 21st century.

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

FUNDING

The funding for this research was supported by the Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang based on the Agreement Letter for Basic Research (Faculty) Funding DPA LPPM UNNES Year 2023 Number 13.7.7/UN37/PPK.04/2023.

REFERENCES

Malik, R. S. (2018). Educational Challenges in 21st Century and Sustainable Development. *Journal of Sustainable Development Education and Research (JSDER)*, *2*(1), 9–20.

ACAR GÜVENDİR, M., & ÖZER ÖZKAN, Y. (2022). Item Removal Strategies Conducted in Exploratory Factor Analysis: A Comparative Study. *International Journal of Assessment Tools in Education*, *9*(1), 165–180. https://doi.org/10.21449/ijate.827950

Afthanorhan, A., Awang, Z., & Aimran, N. (2020). An extensive comparison of cb-sem and pls-sem for reliability and validity. *International Journal of Data and Network Science*, *4*(4), 357–364. https://doi.org/10.5267/j.ijdns.2020.9.003

Álvarez, A., Martín, M., Fernández-Castro, I., & Urretavizcaya, M. (2013). Blending traditional teaching methods with learning environments: Experience, cyclical evaluation process and impact with MAgAdI. *Computers and Education*, *68*, 129–140. https://doi.org/10.1016/j.compedu.2013.05.006

Bahtiar, B., Maimun, M., & Baiq Lily Anggriani W. (2022). Pengaruh Model Discovery Learning Melalui Kegiatan Praktikum IPA Terpadu Terhadap Kemampuan Berpikir Kritis Siswa. *Jurnal Pendidikan Mipa*, *12*(2), 134–142. https://doi.org/10.37630/jpm.v12i2.564

Chen, P., Yang, D., Metwally, A. H. S., Lavonen, J., & Wang, X. (2023). Fostering computational thinking through unplugged activities: A systematic literature review and meta-analysis. *International Journal of STEM Education*, *10*(1). https://doi.org/10.1186/s40594-023-00434-7

Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2023). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. In *Asia Pacific Journal of Management* (Issue 0123456789). Springer US. https://doi.org/10.1007/s10490-023-09871-y

Cholifah, S. N., & Novita, D. (2022). Pengembangan E-LKPD Guided Inquiry-Liveworksheet untuk Meningkatkan Literasi Sains pada Submateri Faktor Laju Reaksi. *Chemistry Education Practice*, *5*(1), 23–34. https://doi.org/10.29303/cep.v5i1.3280

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

Cooper, G. (2023). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology*, *32*(3), 444–452. https://doi.org/10.1007/s10956-023-10039-y

Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.

Dash, G., & Paul, J. (2021). CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting. *Technological Forecasting and Social Change*, *173*(August), 121092. https://doi.org/10.1016/j.techfore.2021.121092

Davy Tsz Kit, N. G., Luo, W., Chan, H. M. Y., & Chu, S. K. W. (2022). Using digital story writing as a pedagogy to develop AI literacy among primary students. *Computers and Education: Artificial Intelligence*, *3*(October 2021), 100054. https://doi.org/10.1016/j.caeai.2022.100054

Dewi, C. A., Khery, Y., & Erna, M. (2019). An ethnoscience study in chemistry learning to develop scientific literacy. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 279–287. https://doi.org/10.15294/jpii.v8i2.19261

Dewi, Citra Ayu, Erna, M., Martini, Haris, I., & Kundera, I. N. (2021). Effect of Contextual Collaborative Learning Based Ethnoscience to Increase Student's Scientific Literacy Ability. *Journal of Turkish Science Education*, *18*(3), 525–541. https://doi.org/10.36681/tused.2021.88

Erduran, S., Ioannidou, O., & Baird, J. A. (2021). The impact of epistemic framing of teaching videos and summative assessments on students' learning of scientific methods. *International Journal of Science Education*, *43*(18), 2885–2910. https://doi.org/10.1080/09500693.2021.1998717

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

Fatiyah, H. N., Riandi, & Solihat, R. (2021). Development of learning tools education for sustainable development (ESD) integrated problem-solving for high school. *Journal of Physics: Conference Series*, *1806*(1). https://doi.org/10.1088/1742-6596/1806/1/012157

Fields, D., Lui, D., Kafai, Y., Jayathirtha, G., Walker, J., & Shaw, M. (2021). Communicating about computational thinking: understanding affordances of portfolios for assessing high school students' computational thinking and participation practices. *Computer Science Education*, *31*(2), 224–258. https://doi.org/10.1080/08993408.2020.1866933

Freeman, L. A., & Jessup, L. M. (2004). The power and benefits of concept mapping: Measuring use, usefulness, ease of use, and satisfaction. *International Journal of Science Education*, *26*(2), 151–169. https://doi.org/10.1080/0950069032000097361

Gebremedhin, M., Gebrewahd, E., & Stafford, L. K. (2022). Validity and reliability study of clinician attitude towards rural health extension program in Ethiopia: exploratory and confirmatory factor analysis. *BMC Health Services Research*, *22*(1), 1–10. https://doi.org/10.1186/s12913-022-08470-9

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. https://doi.org/10.1037/a0032138

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

Han, J., Park, D., Hua, M., & Childs, P. R. N. (2022). Is group work beneficial for producing creative designs in STEM design education? *International Journal of Technology and Design Education*, *32*(5), 2801–2826. https://doi.org/10.1007/s10798-021-09709-y

Heryani, T. P., Suwarma, I. R., & Chandra, D. T. (2023). Development of STEM-Based Physics Module with Self-Regulated Learning to Train Students Critical Thinking Skills. *Jurnal Penelitian Pendidikan IPA*, *9*(6), 4245–4252. https://doi.org/10.29303/jppipa.v9i6.3578

Hettithanthri, U., Hansen, P., & Munasinghe, H. (2023). Exploring the architectural design process assisted in conventional design studio: a systematic literature review. *International Journal of Technology and Design Education*, *33*(5), 1835–1859. https://doi.org/10.1007/s10798-022-09792-9

Hidayat, R., Syed Zamri, S. N. A., & Zulnaidi, H. (2018). Exploratory and confirmatory factor analysis of achievement goals for indonesian students in mathematics education programmes. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(12). https://doi.org/10.29333/ejmste/99173

Hofstein, A., & Mamlok-Naaman, R. (2007). The laboratory in science education: The state of the art. *Chemistry Education Research and Practice*, *8*(2), 105–107. https://doi.org/10.1039/B7RP90003A

INDRAWATI, M. (2017). Keefektifan Lembar Kerja Siswa (Lks) Berbasis Etnosains Pada Materi Bioteknologi Untuk Melatihkan Keterampilan Proses Sains Siswa Kelas Ix. *Pensa: Jurnal Pendidikan Sains*, *5*(02), 152–158.

Isbell, T., & Goomas, D. T. (2014). Computer-assisted rubric evaluation: Enhancing outcomes and assessment quality. *Community College Journal of Research and Practice*, *38*(12), 1193–1197. https://doi.org/10.1080/10668926.2014.899526

Jayanti, M. I., & Nurfathurrahmah, N. (2023). Gerakan Penguatan Literasi Sains Melalui Praktikum Ipa Sederhana Di Smpn 11 Kota Bima. *Taroa: Jurnal Pengabdian Masyarakat*, *2*(1), 1–8. https://doi.org/10.52266/taroa.v2i1.1220

Kanapeckas Métris, K. L. (2020). Activities and assessment solutions for students in advanced molecular genetics and biochemistry to direct and engage with public communication in an online environment. *Biochemistry and Molecular Biology Education*, *48*(5), 439–441. https://doi.org/10.1002/bmb.21389

Kurniawati, A. (2021). Science Process Skills and Its Implementation in the Process of Science Learning Evaluation in Schools. *Journal of Science Education Research*, *5*(2), 16–20. https://doi.org/10.21831/jser.v5i2.44269

Kurukunda, S., Trigona, C., & Baglio, S. (2020). Laboratory Activity during COVID-19 as a "Virtual Experience": Restriction or Chance? *Proceedings of the 17th International Multi-Conference on*

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

Systems, Signals and Devices, SSD 2020, 349–353. https://doi.org/10.1109/SSD49366.2020.9364113

Lawshe, C. H. (1975). A quantitative approach to content validity".Personnel Psychology. *Personnel Psychology*, *28*, 563–575.

Ledesma, R. D., Valero-Mora, P., & Macbeth, G. (2015). The Scree Test and the Number of Factors: a Dynamic Graphics Approach. *The Spanish Journal of Psychology*, *18*, E11. https://doi.org/10.1017/sjp.2015.13

Lee, H., Yoo, J., Choi, K., Kim, S. W., Krajcik, J., Herman, B. C., & Zeidler, D. L. (2013). Socioscientific Issues as a Vehicle for Promoting Character and Values for Global Citizens. *International Journal of Science Education*, *35*(12), 2079–2113. https://doi.org/10.1080/09500693.2012.749546

Malik, R. S. (2018). Educational Challenges in 21St Century and Sustainable Development. *Journal of Sustainable Development Education and Research*, *2*(1), 9. https://doi.org/10.17509/jsder.v2i1.12266

Mang, H. M. A., Chu, H. E., Martin, S. N., & Kim, C. J. (2023). Developing an Evaluation Rubric for Planning and Assessing SSI-Based STEAM Programs in Science Classrooms. *Research in Science Education*, *53*(6), 1119–1144. https://doi.org/10.1007/s11165-023-10123-8

McDermott-Dalton, G. (2022). Putting the 'e' in portfolio design: an intervention research project investigating how design students and faculty might jointly reimagine the design portfolio activity. *International Journal of Technology and Design Education*, *32*(2), 1207–1225. https://doi.org/10.1007/s10798-020-09640-8

Pahrudin, A., Irwandani, Triyana, E., Oktarisa, Y., & Anwar, C. (2019). The analysis of pre-service physics teachers in scientific literacy: Focus on the competence and knowledge aspects. *Jurnal Pendidikan IPA Indonesia*, *8*(1), 52–62. https://doi.org/10.15294/jpii.v8i1.15728

Petrie, C. (2023). Design and use of domain-specific programming platforms: interdisciplinary computational thinking with EarSketch and TunePad. *Computer Science Education*, *00*(00), 1–34. https://doi.org/10.1080/08993408.2023.2240657

Pino, M. E. M., Ordoñez, F. R. R., Ysa, R. A. S., Llanos, D. M. J., Cruz, M. M. T., & Calderón, B. A. C. (2023). Role of Expert in Validation of Information Collection Instruments for Business Purposes. *International Journal of Professional Business Review*, *8*(8), e03122. https://doi.org/10.26668/businessreview/2023.v8i8.3122

Ploj Virtič, M. (2022). Teaching science & technology: components of scientific literacy and insight into the steps of research. *International Journal of Science Education*, *44*(12), 1916–1931. https://doi.org/10.1080/09500693.2022.2105414

Purnamasari, S. (2020). Pengembangan Praktikum IPA Terpadu Tipe Webbed untuk Meningkatkan Keterampilan Proses Sains. *PSEJ (Pancasakti Science Education Journal)*, *5*(2), 8–15. https://doi.org/10.24905/psej.v5i2.20

Pursitasari, I. D., Permanasari, A., Rubini, B., & Ardianto, D. (2023). Pelatihan Penyusunan Desain Praktikum dan Penggunaan KIT Praktikum. *Jurnal Abdinus: Jurnal Pengabdian Nusantara*, *7*(2), 516–530.

Putri*, R. M., Asrizal, A., & Usmeldi, U. (2022). Metaanalisis Efek Pendekatan STEM pada Literasi Sains dan Pemahaman Konsep Peserta Didik di Setiap Satuan Pendidikan. *Jurnal IPA & Pembelajaran IPA*, *6*(1), 86–98. https://doi.org/10.24815/jipi.v6i1.23897

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment and Evaluation in Higher Education*, *35*(4), 435–448. https://doi.org/10.1080/02602930902862859

Rodriguez, S. L., Friedensen, R., Marron, T., & Bartlett, M. (2019). Latina Undergraduate Students in STEM: The Role of Religious Beliefs and STEM Identity. *Journal of College and Character*, *20*(1), 25–46. https://doi.org/10.1080/2194587x.2018.1559198

Rukmini, D., & Saputri, L. A. D. E. (2017). The authentic assessment to measure students' English productive skills based on 2013 Curriculum. *Indonesian Journal of Applied Linguistics*, *7*(2), 263–273. https://doi.org/10.17509/ijal.v7i2.8128

Rusilowati, A., Nugroho, S. E., Susilowati, E. S. M., Mustika, T., Harfiyani, N., & Prabowo, H. T. (2018). The development of scientific literacy assessment to measure student's scientific literacy skills in energy theme. *Journal of Physics: Conference Series*, *983*(1). https://doi.org/10.1088/1742-6596/983/1/012046

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment and Evaluation in Higher Education*, *34*(2), 159–179. https://doi.org/10.1080/02602930801956059

Sari, P. M., & Zulfadewina, Z. (2020). Pengembangan Panduan Praktikum Berbasis Keterampilan Proses Sains Pada Mata Kuliah Praktikum Ipa Sd. *Jurnal Pelita Pendidikan*, *8*(1), 94–98. https://doi.org/10.24114/jpp.v8i1.17334

Setchell, J. M. (2019). Writing a Scientific Report. *Studying Primates*, 271–298. https://doi.org/10.1017/9781108368513.023

Shana, Z., & Abulibdeh, E. S. (2020). Science practical work and its impact on students' science achievement. *Journal of Technology and Science Education*, *10*(2), 199–215. https://doi.org/10.3926/JOTSE.888

Sudaryanto, M., & Akbariski, H. S. (2021). Students' competence in making language skill assessment rubric. *REID (Research and Evaluation in Education)*, *7*(2), 156–167. https://doi.org/10.21831/reid.v7i2.44005

Svenningsson, J., Höst, G., Hultén, M., & Hallström, J. (2022). Students' attitudes toward technology: exploring the relationship among affective, cognitive and behavioral components of the attitude construct. *International Journal of Technology and Design Education*, *32*(3), 1531–1551. https://doi.org/10.1007/s10798-021-09657-7

Timm, J. M., & Barth, M. (2021). Making education for sustainable development happen in elementary schools: the role of teachers. *Environmental Education Research*, *27*(1), 50–66. https://doi.org/10.1080/13504622.2020.1813256

Viñas, L. F. (2022). Testing the Reliability of two Rubrics Used in Official English Certificates for the Assessment of Writing. *Revista Alicantina de Estudios Ingleses*, *36*, 85–109. https://doi.org/10.14198/RAEI.2022.36.05

Wola, B. R., Rungkat, J. A., & Harindah, G. M. D. (2023). Science process skills of prospective science teachers' in practicum activity at the laboratory. *Jurnal Inovasi Pendidikan IPA*, *9*(1), 50–61. https://doi.org/10.21831/jipi.v9i1.52974

Wu, R. M. X., Zhang, Z., Zhang, H., Wang, Y., Shafiabady, N., Yan, W., Gou, J., Gide, E., & Zhang, S. (2023). An FSV analysis approach to verify the robustness of the triple-correlation analysis theoretical framework. *Scientific Reports*, *13*(1), 1–20. https://doi.org/10.1038/s41598-023-35900-3

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409–428. https://doi.org/10.3758/s13428-018-1055-2

ISSN 2406-9205 (print), ISSN 2477-4820 (online)