Journal of Research in Instructional
e-ISSN: 2776-222X
Vol. 6(2) 2026, pp. 907 – 921
https://doi.org/10.30862/jri.v6i2.158

Diagnosis of science process skills of physics education students using three-tier assessment based on scientific reasons

Mahir Mahir*, M. A. Martawijaya, Husain Husain, Abdul Hakim, Abdul Haris

Universitas Negeri Makassar, Indonesia

Submitted: 11-09-2026

Accepted: 07-10-2026

Published: 28-10-2026

Abstract: This study aims to diagnose students' Science Process Skills (SPS) using a three-tier assessment that requires students to provide an answer, select the underlying reason, and indicate their confidence in both responses. The scientific reason refers to the evidence- or principle-based explanation selected by students to justify their answers. The research uses an quantitative approach by involving 15 physics education students selected through nonprobability sampling. The instrument consists of 20 questions representing 10 SPS indicators. The validity of the content of the instrument obtained an Aiken's V coefficient of 0.87. Of the 300 expected responses, 295 responses were filled in and analyzed without imputation. The results showed that the average SPS achievement was 79.11, with the highest achievement the indicator measuring and collecting data of 95.56, while the observing indicator was the lowest achievement of 61.11. The results showed that students were relatively strong in data collection and processing activities, but still needed to strengthen skills related to the early stages of investigation, especially observation, problem formulation, hypothesis formulation, and variable identification. In addition, 16.61% of responses were wrong scientific reasons with high confidence. Diagnostic analysis showed that 57.63% of responses were in the category of mastering SPS, while 19.32% showed pseudo-understanding and 13.22% conceptual errors. The findings showed that three-tier assessments based on scientific reasons can reveal the characteristics of SPS mastery comprehensively that are not seen through the accuracy of answers alone.

Keywords: Diagnostic assessment, scientific reasons, three-tier assessment

Abstrak: Penelitian ini bertujuan untuk mendiagnosis Keterampilan Proses Sains (KPS) siswa menggunakan penilaian tiga tingkat yang mengharuskan siswa memberikan jawaban, memilih alasan yang mendasarinya, dan menunjukkan keyakinan mereka terhadap kedua jawaban tersebut. alasan ilmiah mengacu pada penjelasan berdasarkan bukti atau prinsip yang dipilih oleh siswa untuk membenarkan jawaban mereka. Penelitian menggunakan pendekatan kuantitatif dengan melibatkan 15 mahasiswa pendidikan fisika yang dipilih melalui nonprobability sampling. Instrumen terdiri atas 20 butir soal yang merepresentasikan 10 indikator KPS. Validitas isi instrumen memperoleh koefisien Aiken's V sebesar 0,87. Dari 300 respons yang diharapkan, 295 respons terisi dan dianalisis tanpa imputasi. Hasil penelitian menunjukkan rerata capaian KPS sebesar 79,11, dengan capaian tertinggi pada indicator mengukur dan mengumpulkan data sebesar 95,56, dan terendah pada indikator mengamati sebesar 61,11. Hasil tersebut menunjukkan bahwa mahasiswa relatif kuat pada aktivitas pengumpulan dan pengolahan data, tetapi masih memerlukan penguatan pada keterampilan yang berkaitan dengan tahap awal penyelidikan, khususnya pengamatan, perumusan masalah, perumusan hipotesis, dan identifikasi variabel Sebanyak 82,03% respons merupakan jawaban benar, tetapi hanya 62,37% disertai alasan ilmiah dengan skor maksimal. Selain itu, 16,61% respons merupakan jawaban salah dengan keyakinan tinggi. Analisis diagnostik menunjukkan 57,63% respons berada pada kategori menguasai KPS, sedangkan 19,32% menunjukkan pemahaman semu dan 13,22% kesalahan konseptual. Temuan menunjukkan bahwa three-tier assessment berbasis alasan ilmiah dapat mengungkap karakteristik penguasaan SPS secara komprehensif yang tidak terlihat melalui ketepatan jawaban saja.

Kata kunci: Asesmen diagnostik, alasan ilmiah, asesmen tiga tingkat

This is an open access article under the CC–BY-SA license

*Corresponding author: mahir@unm.ac.id

INTRODUCTION

Science Process Skills (SPS) is one of the important components of scientific competence. It represents an individual's ability to use scientific procedures and reasoning in acquiring evidence-based knowledge. SPS is related to the ability to conduct practicum in physics learning, and includes the process of observing phenomena, identifying problems, formulating problems and hypotheses, determining variables, designing investigations, obtaining and interpreting data, and drawing and communicating conclusions (Ekici & Erdem, 2020; Gizaw & Sota, 2023; Mandasari et al., 2021). This characteristic places SPS as an integrative ability (Kurniawati, 2021), because it connects conceptual knowledge with scientific practice and evidence-based decision-making. For physics education students, mastering SPS has a strategic meaning to complete academic tasks and become part of the competencies that they will transfer to students when carrying out their profession as teachers in the future.

The urgency of SPS development is getting stronger along with the change in the orientation of science learning from learning centered on mastery of facts to learning that emphasizes scientific practice, investigation, argumentation, and the use of evidence, which requires that physics learning outcomes include an understanding of physics concepts and SPS that must be developed in students (Özer & Sarıbaş, 2023; Zembal-Saul, 2009). In line with that, physics education students as prospective teachers need to experience the process of scientific knowledge being built, not just understanding the final product in the form of concepts or laws of physics (Özer & Sarıbaş, 2023; Saribas & Ceyhan, 2015). Learning that provides an experience of the science process allows students to deal with formulating problems, determining relevant information, making predictions, testing conjectures through data, and evaluating the suitability between evidence and conclusions (Darmaji et al., 2019; Maison et al., 2019). Thus, the quality of physics learning at the higher education level is not enough to be reviewed from the ability of students to solve conceptual problems alone, but also from their ability to use scientific processes to produce and evaluate knowledge.

The problem that then arises is how to obtain accurate information about students' mastery of SPS. Previous research examining SPS assessment practices has shown that multiple-choice tests are commonly used to measure students' SPS, although the skills assessed and their representation across test items vary (Lestari, 2023; Rumalolas et al., 2021; Sıbıç & Acar Şeşen, 2022). The use of multiple-choice formats is also supported by their practical advantages, including efficient administration, objective scoring, and the ability to assess large numbers of students within a limited time (Caleon & Subramaniam, 2010). However, the use of conventional multiple-choice items may provide limited information about the reasoning underlying students' responses, particularly when the purpose of assessment is to diagnose the nature of students' understanding rather than merely determine whether their answers are correct.

Tests that base decisions solely on correct or incorrect answers have limitations when used for diagnosis. A correct answer does not necessarily demonstrate scientifically appropriate reasoning because it may result from guessing or other response strategies, whereas an incorrect answer does not necessarily indicate a complete absence of understanding. Caleon and Subramaniam (2010) demonstrated that diagnostic assessment can provide more detailed information by combining students' answers with their reasons

and confidence levels, enabling researchers to distinguish the correctness of an answer from the quality and certainty of the underlying understanding. This distinction is particularly relevant to SPS because these skills involve not only obtaining an answer but also engaging in scientific processes such as identifying variables, formulating hypotheses, interpreting evidence, and drawing conclusions. Research on SPS assessment has likewise indicated that conventional multiple-choice tests may not represent all SPS indicators equally, highlighting the need for assessment approaches that can provide more informative evidence about students' scientific processes (Lestari, 2023; Sıbıç & Acar Şeşen, 2022). Therefore, diagnosing SPS requires assessment information that can distinguish the outcome of a response from the reasoning and confidence underlying that response.

Three-tier assessment is an alternative to answer these limitations through the integration of a number of information in one response (Cheung & Yang, 2020; Irmak et al., 2023; Ma et al., 2025). In this study, the first level was used to obtain the accuracy of the answers, the second level was used to identify students' beliefs, and the third level was to obtain scientific reasons openly. The scientific reason refers to the evidence- or principle-based explanation selected by students to justify their answers. The structure provides an opportunity to identify response patterns that cannot be distinguished through conventional tests, such as correct answers with inadequate reasons, wrong answers with high confidence, or correct answers but students are not yet sure of their responses. By including scientific reasons as part of the assessment, the diagnosis is directed not only to find out if the student is "true", but also to see to what extent the answer can be scientifically accounted for. Students who give complete reasons and according to scientific principles show different mastery characteristics from students who only give correct answers without adequate explanations. Similarly, a partially correct reason can show that students already have certain elements of understanding even though they have not been able to apply them completely. From a formative assessment perspective, this kind of information is more useful because it can provide a final decision on right or wrong and indicate areas of reasoning that need to be improved (Zhu et al., 2020).

Previous research has shown that the three-tier assessment format is widely used to identify students' conceptual understanding, misconceptions, and confidence levels in various areas of science (Sari et al., 2024; Yeo & Yang, 2022). Meanwhile, SPS measurement is more often directed at skill achievement through observation, written tests, laboratory performance, or assessment of investigative activities. This condition shows that there is gap for methodological development, especially the use of three-tier assessment as a diagnostic instrument about students' SPS. This gap is increasingly relevant for physics education students because prospective teachers must have the ability to carry out scientific processes and explain the scientific basis of solving physics problems when becoming teachers in the future. This study takes this position by not treating the three-tier assessment solely as a question format, but as a framework to read the characteristics of student responses. The ten indicators used represent a series of scientific processes, from observing to communicating results. The analysis was then carried out in stages: first, identifying the overall and objective achievement of the SPS, second, analyzing the quality of scientific reasoning, third, looking at the relationship between the accuracy of the answers and the level of confidence, and fourth, mapping the combination of the three information into a diagnostic profile (Betries et al., 2024; Caleon & Subramaniam, 2010).

This approach allows the research to provide a more detailed picture of what students are good at, which parts are not yet strong, and the characteristics of the mistakes that arise.

Previous studies on diagnostic assessment have demonstrated the usefulness of multi-tier formats for identifying students' reasoning and confidence beyond answer accuracy. However, the application of this diagnostic approach to Science Process Skills (SPS) requires further attention, particularly in mapping students' performance across a comprehensive sequence of scientific investigation processes. In the context of physics education, SPS assessment is often reported at the level of individual skills or selected indicators, which may provide limited information about how students perform across the interconnected stages of scientific inquiry. In response to this gap, the present study integrates answer accuracy, scientific reasoning, and confidence into a three-tier diagnostic framework and applies it across ten SPS indicators representing the sequence of scientific investigation, from observing and formulating problems to drawing conclusions and communicating results. The novelty of the study therefore lies not in the use of a three-tier format itself, but in its integration with comprehensive indicator-level mapping of SPS to generate a more differentiated diagnostic profile of physics education students.

Based on these arguments, this research answers four questions: (1) how is the SPS profile of physics education students on the ten indicators measured, (2) What is the quality of the scientific reasons given by students, (3) what is the level of confidence of students in the chosen answers, and (4) How are diagnostic characteristics formed from a combination of the accuracy of answers, beliefs, and scientific reasons. This study aims to diagnose SPS of physics education students using a three-tier assessment based on scientific reasons by mapping the achievement of each indicator and the characteristics of student responses. The results of the research make an empirical contribution to the development of SPS assessments while providing information that can be used to design physics learning that is more oriented to scientific practice and evidence-based reasoning.

METHOD

Research design

This study uses a quantitative approach with an quantitative design (Mertens, 2010). The design was chosen because the research aims to obtain a profile of students' SPS while exploring the characteristics of responses based on the accuracy of answers, confidence levels, and the quality of scientific reasons. Analysis was carried out at the individual level, items, and indicators to obtain a comprehensive picture of SPS achievement and difficulty patterns.

Research subject

The research subjects consisted of 15 physics education students at Makassar State University who were selected using the convenience sample (Sumargo et al., 2024). The subject is a student who has taken four semesters and has completed courses related to inquiry learning. The selection of subjects is based on the suitability of these characteristics with the purpose of the research. The number of subjects is relatively small, so the results of this study are not intended to be generalized statistically to all physics education students. The findings are positioned as an exploratory diagnosis in the context of the

research subject and as a basis for identifying aspects of SPS that can be the focus of research as well as learning interventions and subsequent research.

Research instruments

The research data was collected using a three-tier assessment instrument based on scientific reasons consisting of 20 questions. The instrument was adapted from (Andaria & Hadiwinarto, 2023) and has been validated and tested for reliability before being used in this study. Each indicator is represented by two question items, including observing, formulating problems, formulating hypotheses, identifying variables, defining variables operationally, designing investigations, measuring and collecting data, analyzing data, drawing conclusions, and communicating results. Each indicator is represented by 2 question items, including observing, formulating problems, formulating hypotheses, identifying variables, defining variables operationally, designing investigations, measuring and collecting data, analyzing data, drawing conclusions, and communicating results. Each question item consists of three components. The first component asks students to choose answers from five alternatives. The second component asks students to express their level of confidence in the answers chosen with the choice of "sure" or "unsure". The third component asks students to give reasons openly and scientifically. The validity of the instrument's content was obtained through the assessment of three experts with an Aiken's V coefficient of 0.87, which indicates a good level of expert agreement (Aiken, 1985).

Each item required students to select an answer, indicate their level of confidence, and provide a scientific reason for their answer. Table 1 presents the ten SPS indicators, the corresponding item numbers, and examples of the question contexts used to assess each indicator.

Table 1. SPS indicators and examples of items

Indicator	Item	Example of question context
Observing	1–2	Identifying the direct observation of a falling ball (Item 1) and identifying quantitative observations of temperature changes during water heating (Item 2).
Formulating problems	3–4	Formulating a testable problem concerning the effect of pendulum length on its period (Item 3) and the effect of paper surface area on falling time (Item 4).
Formulating hypotheses	5–6	Formulating a hypothesis about the effect of pendulum length on period (Item 5) and the effect of force on spring extension (Item 6).
Identifying variables	7–8	Identifying the dependent variable in a pendulum experiment (Item 7) and identifying the independent and dependent variables in an electrical-resistance investigation (Item 8).
Operationally defining variables	9–10	Defining solar-panel output using measurable voltage and current (Item 9) and operationally defining dissolution rate based on the time required for sugar to dissolve (Item 10).
Designing investigations	11–12	Designing an investigation of the effect of trolley mass on acceleration while maintaining a constant pulling force (Item

		11) and investigating the effect of wire length while keeping the material and cross-sectional area constant (Item 12).
Measuring and collecting data	13–14	Calculating the period from the time required for 10 oscillations (Item 13) and determining the appropriate value from repeated caliper measurements (Item 14).
Interpreting data	15–16	Interpreting the relationship between force and spring extension (Item 15) and interpreting the relationship between string length and pendulum period (Item 16).
Drawing conclusions	17–18	Drawing a conclusion about the effect of string length on pendulum period (Item 17) and the effect of heater power on water temperature increase (Item 18).
Communicating results	19–20	Selecting an appropriate graph to represent the relationship between heating time and water temperature (Item 19) and formulating an evidence-based statement for a scientific report (Item 20).

Scoring

Correct answers are scored 1 and incorrect answers are scored 0. Scientific reasons are scored 0–2 based on the quality of responses, a score of 2 for reasons that are true, relevant, complete, and demonstrate proper scientific principles, a score of 1 for reasons that are partially correct but incomplete or still common, and a score of 0 for reasons that are incorrect, irrelevant, or do not provide a reason. Thus, each item has a maximum score of 3 and the overall maximum score is 60. Confidence level is not included in the calculation of SPS scores. The variable is used as diagnostic information to interpret the relationship between the accuracy of the answer and the quality of the reason. Out of the 300 expected responses, 295 responses were filled in. Five blank responses came from one student on five items and were not imputed. SPS scores are calculated based on the following equation.

$$\text{SPS} = \frac{\text{Answer Score+Reason Score}}{60} \times 100 \qquad (1)$$

SPS achievement is also calculated based on each indicator, so that relatively strong skills are identified and that require strengthening.

Data analysis

Descriptive analysis was used to obtain the minimum score, maximum score, average, standard deviation, and percentage of SPS achievement (Sugiyono, 2019). The analysis of each indicator is done by comparing the actual score with the maximum score that can be obtained on the two items that represent each indicator. Diagnostic analysis was carried out by integrating three components of response, namely the accuracy of the answer, the level of confidence, and the quality of the reason. A combination of the three components was used to identify the response categories that had been defined in the instrument, including SPS mastery, mastery with low confidence, pseudo-comprehension, possibility of guessing, not yet mastering SPS, conceptual errors, and partial reasoning with incorrect application (Jusniar et al., 2020). Interpretation of conceptual errors is done carefully. Responses to one item are not used singularly to establish misconceptions because such identification

requires consistency on other items that measure the same indicator or confirmation through diagnostic interviews.

RESULTS AND DISCUSSION

Student SPS profile

This study involved 15 physics education students who worked on 20 three-tier assessments based on scientific reasons. Each item contains three components, namely the accuracy of the answers, the level of confidence, and the scientific reasons. Based on the data obtained, there were 300 responses that should have been available from 20 questions by 15 students. However, five responses were not filled out so there were 295 responses that could be analyzed. The five blank responses came from one student in items 7, 9, 11, 19, and 20. Unfilled responses are not imputed.

The SPS score is calculated based on the answer score and reason, while the confidence level is used as diagnostic information. Of the 15 students, the total score was in the range of 21–55 out of a maximum score of 60, with an average of 47.47 and a standard deviation of 8.63. This average is equivalent to an achievement of 79.11. If students with five blank responses were excluded from the sensitivity analysis, the average achievement of 14 students with complete responses was 82.26. These differences show that the existence of empty responses needs to be considered in reading the group's overall achievement.

Table 2. Descriptive statistics of students' science process skills

Statistics	Value
Number of students	15
Theorist maximum score	60
Empirical minimum score	21
Empirical maximum score	55
Average	47.47
Standart Deviation	8.63
Average Reach	79.11

Individually, most students achieve a grade of 75, although there are variations between students. The students with the highest score obtained 55, while the lowest score was 21. The range of achievement shows that the characteristics of student SPS are not completely homogeneous. Thus, the use of group averages alone is not enough to describe individual ability variations.

SPS achievement based on indicators

Analysis of 10 SPS indicators shows a clear difference in achievements. The indicator measuring and collecting data obtained the highest achievement of 95.56, while the observing indicator was the lowest achievement of 61.11. Other low achievements were found in formulating hypotheses (63.33) and identifying variables (64.44).

Table 3 shows a non-linear pattern of achievement. Students are relatively strong in skills related to obtaining and processing data, as demonstrated by high achievement in measuring and collecting data, interpreting data, and drawing conclusions. In contrast, the ability related to the early stages of the investigation, particularly observing, formulating

hypotheses, and identifying variables, showed lower achievement. This pattern suggests that students are better able to carry out or follow up on the investigation process when the information and context of the investigation are available than when they have to determine the basis of the investigation independently. This distinction is important because SPS is a set of interrelated skills. The ability to obtain good data, for example, does not automatically indicate that students have been able to determine the right variables or formulate hypotheses that can be tested.

Table 3. SPS achievement based on indicators

No.	Indicator SPS	Reach
1	Observing	61.11
2	Formulating problems	88.89
3	formulating hypotheses	63.33
4	Identifying variables	64.44
5	defining variables operationally	68.89
6	Planning research	83.33
7	Measure and collect data	95.56
8	Interpreting data	91.11
9	Drawing Conclusions	91.11
10	Communicate	83.33

Quality of scientific reasons

Analysis of scientific reasons showed that out of 295 responses filled, 184 responses (62.37%) obtained a score of 2, 102 responses (34.58%) obtained a score of 1, and 9 responses (3.05%) obtained a score of 0. A score of 2 indicates a true, relevant, complete, and in accordance with scientific principles; a score of 1 indicates a reason that is partially correct but not yet complete; while a score of 0 indicates a false, irrelevant, or not given reason.

Table 4. Quality distribution of scientific reasons

Reason score	Frequency	Percentage
2	184	62.37%
1	102	34.58%
0	9	3.05%
Total	295	100%

The average reason score is 1.59 out of a maximum score of 2, or about 79.66%. The results show that most students have been able to provide relevant reasons, but the quality of scientific justification has not been fully optimal in all responses. More important findings are seen when the quality of the reasons is compared to the accuracy of the answers. Of the 295 responses, there were 242 correct answers (82.03%), while only 184 responses obtained the maximum reason score. Thus, there is a gap of 19.66 percentage points between the correct answer and the reason with maximum quality. These findings show that the accuracy of the answer choices does not necessarily represent the quality of students' reasoning.

Student confidence level

The confidence level showed that 279 responses (94.58%) were in the confident category, 15 responses (5.08%) were unsure, and 1 response (0.34%) was recorded as unsure. When the confidence level was analyzed along with the accuracy of the answers, several combinations of responses were obtained as shown in Table 5.

Table 5. Distribution of the accuracy of answers by confidence level

Response combinations	Frequency	Percentage
True–Sure	230	77.97
True–Unsure	11	3.73%
True–Less Sure	1	0.34%
False–Sure	49	16.61%
False–Unsure	4	1.36%
Total	295	100%

Most of the responses were true with high confidence. However, there were 49 wrong-confident responses (16.61%), while wrong-unsure responses were only 4 (1.36%). This condition shows that there is a discrepancy between beliefs and accuracy in some student responses. This information is not visible if the assessment only records correct or incorrect answers.

Profile diagnostics three-tier assessment

The three response components are then integrated to obtain a diagnostic profile. After the variation in category writing in the data is normalized, seven diagnostic categories emerge.

Table 6. Diagnostic profile of student response

Category diagnostic	Frequency	Percentage
Mastering SPS	170	57.63%
False understanding	57	19.32%
Conceptual errors	39	13.22%
Possible guessing	9	3.05%
Mastering but not yet/unsure	8	2.71%
Not yet mastering SPS	6	2.03%
Partial but misapplied reasoning	6	2.03%
Total	295	100%

The category of mastering SPS was the most dominant category with 170 responses or 57.63%. However, as many as 42.37% of responses were outside the mastery category. This proportion mainly came from the false-understanding category of 19.32% and conceptual errors of 13.22%. These results show that measurements based on total scores alone can hide the characteristics of student responses. Students can give the correct answer, but the accompanying reasons do not show strong mastery. In contrast, incorrect answers also have different characteristics depending on the level of belief and the quality of the reason.

The average SPS achievement of 79.11 shows that in general students have relatively good ability to complete tasks that represent the scientific process. In contrast, lower

achievement in observing, formulating hypotheses, and identifying variables indicates difficulties in the early stages of the investigation. All three skills require the ability to recognize relevant information, determine testable relationships, and make decisions about the components of the investigation. In other words, students are relatively stronger in conducting and following up on investigations than in building an investigation structure from the beginning. These findings have consequences for the education of prospective teachers. An overly structured practicum experience can provide inadequate experience to formulate problems, hypotheses, and variables independently (Chiu et al., 2021).

The difference between 82.03% of correct answers and 62.37% of reasons with maximum scores is an important finding in this study. This difference shows that students are more often able to determine the right alternative answer than to provide a complete scientific explanation to support the choice. This finding shows the limitations of interpretation if the assessment only uses answer scores (McNeish & Wolf, 2020). If only the answers were counted, all 242 correct responses would be treated as equivalent proof of mastery. In fact, the analysis of students' reasons showed that some correct responses were not accompanied by adequate scientific justification. This finding is consistent with previous research showing that a correct answer does not necessarily indicate complete or scientifically coherent understanding, as students may select the correct response without demonstrating the underlying reasoning required to support it (Haudek et al., 2012; Sands et al., 2018).

The proportion of reasons for a score of 1 of 34.58% also provides important information. The response cannot be seen as the same as a completely wrong reason because students still show some relevant elements of reasoning. This condition can be interpreted as an incomplete understanding or the ability to apply principles that are not yet consistent. From a learning perspective, such information is more useful than simply knowing that the student's answer is correct. Thus, scientific reason serves as a window into the student's reasoning process (Cabello et al., 2021; Masters, 2020; Watkins & Manz, 2022). This component allows educators to see the basis of thought that underlies the choice of answers. For prospective physics teachers, the ability to give reasons also has pedagogical value because teachers need to be able to explain the relationship between phenomena, principles of physics, evidence, and conclusions in a sequential manner.

The finding that 16.61% of responses were incorrect answers provided important diagnostic information. Students in this category not only gave incorrect responses, but also had confidence that the responses were correct. This condition is pedagogically different from students who answered incorrectly but stated that they were unsure. Incorrect answers with low confidence can indicate an awareness of limited understanding (Grazziotin-Soares et al., 2021; Hull et al., 2022). On the other hand, incorrect answers with high confidence indicate that students are not aware of problems in their thinking process. Therefore, students in both conditions should not receive the same learning treatment. However, false convictions cannot be directly equated with misconceptions. Mistakes can be caused by various factors, including errors in the application of principles, interpretation of information, or incomplete reasoning. In this study, the use of the term conceptual error is more appropriate to describe the data obtained.

The diagnostic profile showed that 19.32% of the responses were in the category of pseudo-understanding. This proportion is large enough to be of concern because it shows

that there are responses that appear to be correct but have not been fully supported by adequate scientific reasoning. In conventional assessments, the response will be counted as correct and contribute to an increase in scores. As a result, educators can conclude that students have mastered certain skills, even though when asked to explain the reason, the quality of the response has not supported the conclusion. Three-tier assessments allow these differences to be seen. Pedagogically, students with pseudo-understanding are not in the same condition as students who have not really mastered SPS. They already have the ability to produce the right answers, but do not have adequate justification.

Low achievement in the indicators observes phenomena, formulates hypotheses, and identifies variables to form a fairly consistent pattern. These three are skills that are in the early stages of the investigation process and determine the direction of the next activity. Difficulties at this stage can cause students to be able to conduct experiments procedurally but do not fully understand the basis of decision-making in the experiment (Kranz et al., 2023; Ploj Virtič, 2022). The indicator identified variables only reached 64.44%. These findings show that students are not completely consistent in determining the aspects that must be manipulated, observed, or controlled in an investigation. In fact, the identification of variables is the basis for determining the relationships to be tested and the type of data that needs to be collected (Lim, 2025; Park et al., 2020).

Similarly, the ability to formulate a hypothesis, which only reached 63.33%, shows that students still have difficulty in turning a problem into a testable prediction. A scientific hypothesis is an alleged answer that needs to show a reasonable relationship between variables based on initial knowledge or available evidence (Ghasemi et al., 2025). The findings suggest that SPS learning needs to provide more experience at the investigation planning stage. Students need to be given the opportunity to determine research questions for themselves, formulate hypotheses, choose variables, and explain the reasons for choosing the investigation design. This approach allows for the development of an early-stage SPS that is the basis for data collection and interpretation skills.

The high achievement in the indicators of data interpretation and conclusion drawn, 91.11%, respectively, also shows that students are quite capable of using data as a basis for decision-making. The findings of this study have direct implications for inquiry learning in physics education. If students are relatively strong in data collection and interpretation but weaker in hypothesis formulation and variable identification, then learning should not only provide experience in conducting experiments whose procedures have been determined.

Inquiry activities can be designed with a higher level of openness, for example, students are given a phenomenon or problem and then asked to formulate research questions, propose hypotheses, determine variables, and design ways to obtain evidence (Baur & Emden, 2021; Wu et al., 2021). In the process, students also need to be asked to provide reasons for each decision made. The strategy directly targets weaknesses found through diagnosis. The results of the study also show that assessments can be used as a basis for providing different feedback. Students who demonstrate pseudo-understanding need argumentation and justification activities, while students with conceptual errors need the opportunity to re-evaluate the relationship between concepts and evidence. Students who are correct but less confident can be directed to reflect and verify the thought process.

The findings of the study showed that the three-tier assessment provided more detailed information about SPS than if the analysis was based only on true-false answers.

The added value is clearly seen from the three findings: 82.03% of the answers are correct, but only 62.37% of reasons achieve the maximum score; 16.61% of the responses are wrong; and 19.32% of the responses are in the category of pseudo-understanding. The three findings show that the SPS achievement has a layer of information that cannot be captured by the answer score alone. The answers provide information about the results of the responses, the beliefs provide information about the students' certainty of the responses, while the reasons provide information about the basis of reasoning. The integration of the three results results in a more informative profile for the purpose of diagnosis.

CONCLUSION

This study shows that the SPS profile of physics education students cannot be fully represented by answer accuracy alone. The findings confirm that integrating answer accuracy, beliefs, and scientific reasons provides information that is not visible when assessment relies only on true-false scores. Pedagogically, the diagnostic results can guide reinforcement according to student response characteristics, particularly through inquiry learning, which provides space to formulate hypotheses, determine variables, design investigations, and justify evidence-based decisions.

The contribution of this research does not lie in claiming that three-tier assessment is a new format, but in its application to diagnose SPS characteristics through scientific reasons. This approach enables educators to distinguish students who have truly mastered the skills from those who obtain correct answers without strong reasoning. The findings should be interpreted within the sample characteristics. The study involved 15 students selected through nonprobability sampling; therefore, the results are not intended for statistical generalization to all physics education students, but as exploratory diagnoses of ability patterns in the studied group. Each indicator was represented by only two items, so indicator-level diagnoses remain preliminary. Future research can increase items per indicator and involve students from several institutions to test diagnostic stability. Interviews or performance observations can strengthen interpretations of categories such as pseudo-understanding and conceptual errors. Despite these limitations, the findings demonstrate that SPS measurement needs to consider response quality more deeply. Diagnosis based on answer accuracy, beliefs, and reasons can serve as a starting point for developing assessments more sensitive to students' reasoning processes.

ACKNOWLEDGEMENT

This research was funded by the Institute for Research and Community Service (LPPM) of the State University of Makassar through Postgraduate PNBP funding with a contract number 507/DST/UN36.12/TU/2026.

REFERENCES

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Andaria, M., & Hadiwinarto, H. (2023). Development and Validation of a Two-Tier Multiple Choice Instrument to Assess Students' Science Process Skills in Acid-Base Chemistry. *ISEJ : Indonesian Science Education Journal,* *4*(1), 11–18.

https://doi.org/10.62159/isej.v4i1.141
Baur, A. & Emden, M. (2021). How to Open Inquiry Teaching? An Alternative Teaching Scaffold to Foster Students' Inquiry Skills. *Chemistry Teacher International, 3*(1), 20190013. https://doi.org/10.1515/cti-2019-0013
Betries, M., Leoniza, T., Styawati, R., Umami, S. R., Thohir, M. A., & Utama, C. (2024). Identifikasi Miskonsepsi Materi Sistem Tata Surya Menggunakan Three-Tier Diagnostic Test dengan Certainty Response Index. *Pendipa: Journal of Science Education, 9*(1), 1–9. https://doi.org/10.33369/pendipa.9.1.1-9
Cabello, V. M., Moreira, P. M., & Griñó, M. P. (2021). Elementary Students' Reasoning in Drawn Explanations Based on a Scientific Theory. *Education Sciences, 11(*10), 581. https://doi.org/10.3390/educsci11100581
Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, *40*(3), 313–337. https://doi.org/10.1007/s11165-009-9122-4
Caleon, I., & Subramaniam, R. (2010). Development and Application of a Three-Tier Diagnostic Test to Assess Secondary Students' Understanding of Waves. *International Journal of Science Education, 32*(7), 939–961. https://doi.org/10.1080/09500690902890130
Cheung, K. L., & Yang, D.-C. (2020). Performance of Sixth Graders in Hong Kong on a Number Sense Three-Tier Test. *Educational Studies*, *46*(1), 39–55. https://doi.org/10.1080/03055698.2018.1516631
Chiu, T. K. F., Chai, C. S., Williams, P. J., & Lin, T. J. (2021). Teacher Professional Development on Self-Determination Theory–Based Design Thinking in STEM Education. *Educational Technology & Society*, *24*(4), 153–165. https://www.jstor.org/stable/48629252
Darmaji, D., Kurniawan, D. A., & Irdianti, I. (2019). Physics Education Students' Science Process Skills. *International Journal of Evaluation and Research in Education*, *8*(2), 293–298. https://doi.org/10.11591/ijere.v8i2.16401
Ekici, M., & Erdem, M. (2020). Developing Science Process Skills through Mobile Scientific Inquiry. *Thinking Skills and Creativity*, *36*. https://doi.org/10.1016/j.tsc.2020.100658
Ghasemi, A., Hosseinpanah, F., Kashfi, K., & Bahadoran, Z. (2025). Research Hypothesis: A Brief History, Central Role in Scientific Inquiry, and Characteristics. *Addiction & Health, 17*, 1623. https://doi.org/10.34172/ahj.1623
Gizaw, G. G., & Sota, S. S. (2023). Improving Science Process Skills of Students: A Review of Literature. *Science Education International*, *34*(3), 216–224. https://doi.org/10.33828/sei.v34.i3.5
Grazziotin-Soares, R., Blue, C., Feraro, R., Tochor, K., Ardenghi, T. M., Curtis, D., & Ardenghi, D. M. (2021). The Interrelationship Between Confidence and Correctness in a Multiple-Choice Assessment: Pointing Out Misconceptions and Assuring Valuable Questions. *BDJ Open*, *7*(1), 10. https://doi.org/10.1038/s41405-021-00067-4
Haudek, K. C., Prevost, L. B., Moscarella, R. A., Merrill, J., & Urban-Lurain, M. (2012). What are they thinking? Automated Analysis of Student Writing About Acid-base Chemistry in Introductory Biology. *CBE Life Sciences Education*, *11*(3), 283–293.

https://doi.org/10.1187/cbe.11-08-0084

Hull, M. M., Jansky, A., & Hopf, M. (2022). Does Confidence in a Wrong Answer Imply a Misconception? *Physical Review Physics Education Research 18*(2), 20108. https://doi.org/10.1103/PhysRevPhysEducRes.18.020108

Irmak, M., Inaltun, H., Ercan-Dursun, J., Yanis-Kalleci, H., & Yuruk, N. (2023). Development and Application of a Three-Tier Diagnostic Test to Assess Pre-service Science Teachers' Understanding on Work-Power and Energy Concepts. *International Journal of Science and Mathematics Education*, *21*(1), 159–185. https://doi.org/10.1007/s10763-021-10242-6

Jusniar, J., Effendy, E., Budiasih, E., & Sutrisno, S. (2020). Developing a Three-Tier Diagnostic Instrument on Chemical Equilibrium (TT-DICE). *Educación Química*, *31*, 84–102. https://dx.doi.org/10.22201/fq.18708404e.2020.3.72133

Kranz, J., Baur, A., & Möller, A. (2023). Learners' Challenges in Understanding and Performing Experiments: A Systematic Review of the Literature. *Studies in Science Education*, *59*(2), 321–367. https://doi.org/10.1080/03057267.2022.2138151

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

Lestari, T. (2023). An Analysis of Junior High Student's Science Process Skills After Covid-19. *Magister Scientae*, *51*(1), 31–38. https://doi.org/10.33508/mgs.v51i1.4501

Lim, W. M. (2025). What Is Quantitative Research? An Overview and Guidelines. *Australasian Marketing Journal*, *33*(3), 325–348. https://doi.org/10.1177/14413582241264622

Ma, H., Yang, H., Li, C., Ma, S., & Li, G. (2025). The Effectiveness and Sustainability of Tier Diagnostic Technologies for Misconception Detection in Science Education: A Systematic Review. *Sustainability*, *17*(7), 3145. https://doi.org/10.3390/su17073145

Maison, M., Darmaji, D., Kurniawan, D. A., Astalini, A., Dewi, U. P., & Kartina, L. (2019). Analysis of Science Process Skills in Physics Education Students. *Jurnal Penelitian dan Evaluasi Pendidikan*, *23*(2), 197–205. https://doi.org/10.21831/pep.v23i2.28123

Mandasari, F., Iwan, I., & Damopolii, I. (2021). The relationship between science process skills and biology learning outcome. *Journal of Research in Instructional*, *1*(1), 23–32. https://doi.org/10.30862/jri.v1i1.9

Masters, H. (2020). Using Teaching Rehearsals to Prepare Preservice Teachers for Explanation-Driven Science Instruction. *Journal of Science Teacher Education*, *31*(4), 414–434. https://doi.org/10.1080/1046560X.2020.1712047

McNeish, D., & Wolf, M. G. (2020). Thinking twice About Sum Scores. *Behavior Research Methods*, *52*(6), 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

Mertens, D. M. (2010). *Research and Evaluation in Education and Psychology Integrating Diversity With Quantitative, Qualitative, and Mixed Methods*. SAGE Publications.

Özer, F., & Sarıbaş, D. (2023). Exploring Pre-service Science Teachers' Understanding of Scientific Inquiry and Scientific Practices Through a Laboratory Course. *Science & Education*, *32*(3), 787–820. https://doi.org/10.1007/s11191-022-00325-3

Park, Y. S., Konge, L., & Artino Jr, A. R. (2020). The Positivism Paradigm of Research.

Academic Medicine, *95*(5), 690–694. https://doi.org/10.1097/ACM.0000000000003093

Ploj, V. 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

Rumalolas, N., Rosely, M. S., Nunaki, J. H., Damopolii, I., & Kandowangko, N. Y. (2021). The inquiry-based student book integrated with local resources: The impact on student science process skill. *Journal of Research in Instructional*, *1*(2), 133–146. https://doi.org/https://doi.org/10.30862/jri.v1i2.17

Sands, D., Parker, M., Hedgeland, H., Jordan, S., & Galloway, R. (2018). Using Concept Inventories to Measure Understanding. *Higher Education Pedagogies*, *3*(1), 173–182. https://doi.org/10.1080/23752696.2018.1433546

Sari, D. N., Arif, K., Yurnetti, & Putri, A. N. (2024). Identification of Students ' Misconceptions in Junior High Schools Accredited A using the Three Tier Test Instrument in Science Learning. *Jurnal Penelitian Pendidikan IPA*, *10*(1), 1–11. https://doi.org/10.29303/jppipa.v10i1.5064

Saribas, D., & Ceyhan, G. D. (2015). Learning to Teach Scientific Practices: Pedagogical Decisions and Reflections During a Course for Pre-service Science Teachers. *International Journal of STEM Education*, *2*(1), 7. https://doi.org/10.1186/s40594-015-0023-y

Sıbıç, O., & Acar, S. B. (2022). Examining Science Process Skills Tests: A Case of Turkey TT - 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

Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

Sumargo, B., Budyanra, & Kurniawan, R. (2024). *Metode dan Pengaplikasian Teknik Sampling*. Bumi Aksara.

Watkins, J., & Manz, E. (2022). Characterizing Pedagogical Decision Points in Sense-Making Conversations Motivated by Scientific Uncertainty. *Science Education*, *106*(6), 1408–1441. https://doi.org/https://doi.org/10.1002/sce.21747

Wu, X. Ben, Sandoval, C., Knight, S., Jaime, X., Macik, M., & Schielack, J. F. (2021). Web-based Authentic Inquiry Experiences in Large Introductory Classes Consistently Associated with Significant Learning Gains for All Students. *International Journal of STEM Education*, *8*(1), 31. https://doi.org/10.1186/s40594-021-00290-3

Yeo, J., & Yang, H. (2022). Using A Three-Tier Multiple- Choice Diagnostic Instrument Toward Alternative Conceptions Among Lower-Secondary School Students in Taiwan: Taking Ecosystems Unit as An Example. *Journal of Baltic Science Education*, *21*(1), 69–83. https://doi.org/10.33225/jbse/22.21.69

Zembal-Saul, C. (2009). Learning to Teach Elementary School Science as Argument. *Science Education*, *93*(4), 687–719. https://doi.org/https://doi.org/10.1002/sce.20325

Zhu, M., Liu, O. L., & Lee, H.-S. (2020). The Effect of Automated Feedback on Revision Behavior and Learning Gains in Formative Assessment of Scientific Argument Writing. *Computers & Education*, *143*, 103668. https://doi.org/https://doi.org/10.1016/j.compedu.2019.103668