Development of a Science Process Skills (SPS) test for assessing environmental chemistry concept: Salt hydrolysis in grade XI

Eka Puspita Kartika Sari[1]*, Muntholib[2], Umi Husnun Nihayah[3]

[1]Departement of Science Education, Faculty of Education, Universitas Jember, Kalimantan Street No.37, Jember, 68121, Indonesia

[2]Department of Chemistry Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, Semarang Street No. 5, Malang, 65145, Indonesia

[3]Department of Chemistry, School of Molecular and life Science, Curtin University, Perth, Australia

*Corresponding author, email: 1997042120224062001@mail.unej.ac.id

Article History
Received: 27 July 2025
Revised: 26 August 2025
Accepted: 27 August 2025
Published: 1 September 2025

Keywords
Instrument
Salt Hydrolysis
Science Process Skills

Abstract
Science process skills (SPS) are fundamental to inquiry-based science learning and important for linking chemistry concepts with environmental issues. However, valid instruments to assess SPS in specific topics, such as salt hydrolysis, are still limited. This study aimed to develop and validate a multiple-choice test to evaluate eleventh-grade students' SPS in the context of salt hydrolysis, emphasizing environmental chemistry applications. A research and development (R&D) design with five stages was employed: literature review, item development, expert validation, pilot testing, and finalization. Four chemistry education specialists provided expert judgment, and a pilot study was conducted with 150 eleventh-grade students from five high schools. Item validity, difficulty, discrimination index, and reliability were analyzed. The final test contained 28 items with a high content validity index (CVI = 0.90). Item analysis indicated two easy, 25 medium, and one difficult item. Discrimination indices classified five items as very good, 20 good, two sufficient, and one poor. All items were empirically valid, and the reliability coefficient (Cronbach's alpha = 0.891) demonstrated strong internal consistency. Compared with similar SPS instruments, this test integrated authentic environmental contexts, such as the impact of salt hydrolysis on water pH, thereby enhancing ecological validity. The instrument provides teachers with a reliable diagnostic tool to identify students' strengths and weaknesses in SPS. Beyond assessment, the instrument supports formative and summative evaluations, integrates environmental issues into chemistry learning, and serves as a reference for developing similar instruments in other science domains.

How to cite: Sari, E. P. K., Muntholib, M., & Nihayah, U. H. (2025). Development of a Science Process Skills (SPS) test for assessing environmental chemistry concept: Salt hydrolysis in grade XI. *Journal of Environment and Sustainability Education, 3*(3). 350–359. doi: 10.62672/joease.v3i3.117

1. Introduction

Science education now faces an increasing demand to not only equip students with knowledge of the discipline, but also with the ability to apply scientific methods to solve real-world problems. Central to this vision are Science Process Skills (SPS), widely recognized as the operational dimension of scientific literacy (Carey et al., 1989; Harlen, 1999). SPS represent the fundamental actions, reasoning processes, and epistemic practices that enable students to observe, hypothesize, design experiments, interpret data, and evaluate evidence. In essence, they are the "tools of the scientist" and are indispensable for inquiry-based science education (Semiawan et al., 1992; Tosun, 2019).

Recent empirical studies have shown that students with strong science process skills (SPS) tend to have a deeper conceptual understanding, enhanced problem-solving abilities, and better critical reasoning skills (Danczak et al., 2020). Beyond individual academic outcomes, systematically promoting SPS has been shown to produce scientifically literate citizens who can critically engage with urgent global issues such as sustainability transitions, public health crises, and rapid technological innovations(Lazonder & Janssen, 2021; Liu et al., 2023). Therefore, SPS should be regarded as essential societal competencies that prepare future generations for meaningful participation in knowledge-driven economies and evidence-informed democratic decision-making, rather than merely instructional targets (Mede et al., 2025).

Despite the broad consensus about their importance, measuring and fostering SPS remains a persistent challenge. One difficulty lies in the construct's breadth: SPS encompasses basic skills such as observation and

doi: 10.62672/joease.v3i3.117

ISSN: 3025-0714

classification, core competencies like hypothesizing and controlling variables, and extended processes like modeling and evidence-based reasoning. Traditional assessments often reduce this complexity to a narrow subset of tasks, which underrepresents higher-order skills (Godec, & King, 2021; Park, & Song, 2021). Another challenge is interpretation validity: many instruments rely on face validity and expert review, yet lack deeper forms of evidence, such as response-process validation, structural analysis, and fairness testing across groups (Crandell, 2024; Yan & Pastore, 2022).

Recent studies have advanced this field by emphasizing rigorous validation practices. For instance, Danczak (2020) developed and validated a critical thinking assessment in chemistry through expert review, pilot testing, and psychometric modeling. Grieger (2022) produced an instrument that measures knowledge of green chemistry principles (ASK-GCP), integrating factor analysis and evidence of criterion-related validity. Lazonder & Janssen (2021) designed a performance-based scientific reasoning test that provided strong validity evidence using Rasch modeling. Similarly, Mi et al. (2023) validated a four-tier diagnostic inventory in physics, and (Liu et al., 2023) examined the fairness and dimensionality of a new STEM identity instrument. These examples highlight the trend toward comprehensive, multi-source validation strategies in science education.

Nevertheless, systematic reviews indicate significant gaps. For example, Yan & Pastore (2022) found that many educational assessments neglect response-process data and differential item functioning (DIF) analyses. This limits confidence in the fairness of scores. These critiques align with the broader movement toward open science and reproducibility in educational research. This movement emphasizes the importance of detailed reporting of analytic decisions, psychometric indices, and limitations (Ye, L. et al., 2023).

In light of these challenges, scholars have proposed organizing SPS into hierarchical frameworks that reflect developmental progression. Building upon classical theory and recent empirical work, this study adopts a three-level SPS model: Basic (B-SPS), Causal (C-SPS), and Experimental (E-SPS). B-SPS includes observing, classifying, measuring, communicating, and recording data. The C-SPS involve inferring, predicting, defining operationally and identifying variables. The E-SPS extend to making hypotheses, designing experiments, changing and controlling variables, modeling and interpreting data (Carey et al., 1989; Tosun, 2019). This model clarifies conceptual boundaries and provides diagnostic granularity. Teachers can identify which foundational practices, intermediate reasoning, or advanced modeling students struggle with. Meanwhile, researchers can investigate how skills develop across levels and contexts. The comprehensive SPS framework used in this study can be seen in Table 1.

Table 1. process skill framework

Domain	Component Science Process Skills	Descriptors
B-SPS	Observation	The process of obtaining information/data using the senses (eyes, ears, nose, skin, tongue); gathering and using relevant facts.
	Classifying	Find differences, similarities, compare, group, or connect existing data.
	Measurements	Performing measurements related to known data
	Communication	Describe empirical data from observations or experiments using graphs or tables; compile reports, explain experimental results, discuss the results of activities that have been carried out; read graphs, tables, or diagrams.
	Data recording	Record the results of the experiment and explain the data found.
C-SPS	Drawing conclusion	Drawing conclusions or conclusions from experiments or facts
	Predicting	Express what might happen in the situation described in the question.
	Defining Operationally	Defining a concept to make it measurable
	Identify variables	Determining the variables used in specific situations
E-SPS	Creating a hypothesis	Determining the possibilities of a given situation
	Designing an experiment	Designing an experiment based on the description
	Changing and controlling variables	Determining variables from a modified experiment
	Modeling	Modeling experiments in a simpler way
	Data interpretation	Connecting observations, identifying patterns in a series of observations

The topic of salt hydrolysis has domains that are highly relevant to the components of the science process skills. Based on the analysis results, salt hydrolysis has phenomena that can be observed. It also has types that can be classified. Furthermore, it has quantities that can be measured. In addition, it has facts, concepts, and procedures that can be communicated and defined operationally. It also has data from observations and measurements that can be analyzed and interpreted. Finally, it has patterns that can be concluded and used to predict. It has research procedures that identify variables, which are then made into hypotheses. It has steps, materials, and tools used for experiments, which can be used to design an experiment. It has properties relevant to the components of science process skills. It has properties that can be used to change and control variables. It has several experiments that can be modeled. This means that the science of salt hydrolysis can be used as material in developing science process skills assessment instruments to assess and develop students' science process skills.

Developing instruments that align with this framework requires adherence to rigorous psychometric protocols. First, domain specification and blueprinting ensure alignment between construct definitions and item targets. Second, expert judgment verifies clarity and minimizes irrelevant variance, such as linguistic complexity (Deng et al., 2021). Third, response-process validation through think-alouds confirms that students interpret items as intended. Fourth, pilot testing with Rasch or IRT modeling examines dimensionality, threshold ordering, and fairness across subgroups (Ye, L. et al., 2023; Zhang et al., 2020). Finally, the interpretation and use phases rely on Wright maps and proficiency bands to link scores to actionable instructional decisions (Park, M., & Song, 2021).

However, even with these sophisticated models, SPS assessments must address emerging challenges. One such challenge is equity. As Ye, L. et al. (2023) demonstrated, ignoring differential item functioning (DIF) can introduce biases in interpretations based on gender or school type. Another challenge is the rise of digital and hybrid learning environments, which complicate comparability across modes of administration. Additionally, the transferability of SPS across disciplines—for instance, from chemistry to environmental science—is still an open empirical question (González-Cabrera et al., 2021; Zhou & Lee, 2021).

In response, researchers have emphasized iterative cycles of development and validation that incorporate multiple forms of evidence across contexts and populations (Ding et al., 2022; Zhang et al., 2020). This approach aligns with international best practices in educational measurement, which view validation as an ongoing process rather than a one-time procedure (Yan & Pastore, 2022). The present study integrates insights from classical foundations(Carey et al., 1989; Harlen, 1999; Semiawan et al., 1992) and contemporary advances (Danczak et al., 2020; Grieger et al., 2022; Mi et al., 2023) to contribute an SPS instrument that is theoretically grounded and psychometrically robust.

The purpose of this research and development is to produce a multiple-choice science process skills test on grade 11 Salt Hydrolysis material consisting of questions that are valid and have reliability that meets the requirements for learning and research purposes.

2. Method

This study used a research and development (R&D) approach to create an assessment instrument in the form of a multiple-choice test on integrated science process skills (SPS) related to salt hydrolysis for eleventh grade SMA students. The instrument development process consists of five stages (Figure 1): (1) a literature review, (2) developing question items, (3) an expert assessment, (4) a field trial, and (5) finalizing the instrument (Damanhuri et al., 2016; Muntholib, et al., 2020a; Muntholib, et al., 2020b; Muntholib, et al., 2020c).

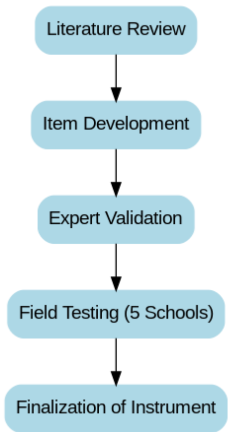

Figure 1. Research flowchart of SPS test development

During the literature review stage, we examined the learning objectives for Grade 11 salt hydrolysis in accordance with the relevant curriculum (MoEC, 2016), the science process skills framework (Tosun, 2019), and the subject matter of salt hydrolysis (Chang & Goldsby, 2016; Lewis & Evans, 2018; McMurry et al., 2015). We then identified pertinent issues. This stage produced an appropriate context problem map.

During the item development stage, we compiled a grid of questions, identified possible questions, selected the most appropriate ones, and developed the items. At this stage, we produced 28 items involving various kinds of salt hydrolysis, the properties of hydrolyzed salt, and pH calculation. We used examples from daily life and followed the framework of science process skills.

The expert validation stage involved four experienced experts. The experts were: (1) two university lecturers with doctoral degrees in chemistry education and over 10 years of teaching and research experience; (2) a high school chemistry teacher with a master's degree in science education and national teaching recognition; and (3) a chemistry curriculum specialist with a master's degree and 10 years of curriculum development experience. This diversity of experts provides stronger evidence of the instrument's validity. The experts assessed each item with the criteria of (1) the suitability of the item with the coverage of the subject matter, (2) the suitability of the item with the framework of science process skills, (3) the suitability of each item with its indicators, (4) the truth of knowledge, and (5) the clarity of language. The assessment of each item was carried out with five scales, namely: 1 if only one criterion is met; 2 if only two criteria are met; 3 if only three criteria are met; 4 if only four criteria are met; and 5 if all criteria are met. Validity can be calculated using CVI which is useful for calculating how valid the instrument developed is. The CVI criteria are poor (B) if <0.00; slightly or unacceptable (D) if 0.00-0.20; fair (C) or questionable if 0.21-0.40; moderate or good (S) if 0.41-0.60; quite good or very good (CB) if 0.61- 0.80; almost perfect or excellent (HS) if 0.81-1.00 (Landis & Koch, 1977; Regier et al., 2013).

The field trial phase of the Science Process Skills Test took place in Banyuwangi Regency. The trial involved 150 students from five high schools. The schools were purposely selected based on their accreditation status (A level), adequate laboratory facilities, and experienced chemistry teachers who regularly teach hydrolysis concepts. Thirty students were selected from each school to ensure proportional representation and capture variation between urban and semi-urban schools. Purposive sampling is considered appropriate in test development studies to ensure the representativeness of the target population.

Ethical approval for this study was obtained from the Research Ethics Committee of Universitas Negeri Malang. Written informed consent was obtained from the students and their guardians, and permission was secured from the respective school principals. All procedures adhered to research ethics guidelines for studies involving minors and followed international standards for educational research ethics (Pluye, P., et al. 2021)

Distinguishing power reflects the ability of the question to distinguish students who are able from those who are less able to solve the problem. The question's differentiation power (D) is determined by determining the difference between the difficulty index of the upper group students (Pa) and the lower group (Pb); D= Pa - Pb. Upper group students (Pa) are students who obtain the highest test score (30%), while lower group students (Pb) are students who obtain the lowest test score (30%). The differential power of a question is categorized as poor if $0.00< D\leq 0.20$; sufficient if $0.21 < D\leq 0.40$; good if $0.41 < D\leq 0.70$; and very good if $0.71< D \leq 1.00$.

The difficulty index is a number that fulfills the difficulty level of a problem to be solved. The item difficulty index (P) is calculated by comparing the number of students who answered correctly (B) with the number of students who took the test (Js); P= B/Js. Question items are categorized as difficult if $0\leq P\leq 0.3$; medium when $0.3< P\leq 0.7$; and easy when $0.7< P\leq 1$.

Item validity measures the correlation between the item and the assessment instrument score. An item is declared valid if the respondent who answers the question correctly has a high test score. The acceptable value of the biserial point coefficient of an item is 0.2 (Ding & Beichner, 2009; Wuttiprom et al., 2009). Item validity was calculated using the product-moment correlation formula conducted with the SPSS 22.0 for Windows program.

Reliability is the extent to which an instrument provides the same measurement results when measurements are repeated (Taber, 2018). When used to measure the same object, a reliable instrument will give the same or almost the same results even though measurements are taken at different times (Kimberlin & Winterstein, 2008). However, some science education journals offer a Chronbach's alpha coefficient of 0.7 as an acceptable value, although others offer lower values (van-Griethuijsen et al., 2015).

Finalization was carried out to determine the final form of the salt hydrolysis material science process skills test. Finalization considers the following: (1) Test appearance; (2) Choice of sentence structure; (3) Representation of salt hydrolysis subject matter in accordance with curriculum demands; (4) Representation of knowledge and skills in accordance with the established science process skills framework; (5) Representation of the level of difficulty of the items

3. Results and Discussion

The present study aimed to develop and validate a Science Process Skills (SPS) test on salt hydrolysis for eleventh-grade students. Twenty-eight items were constructed and evaluated based on content validity, discrimination, difficulty level, empirical validity, and reliability. The findings suggest that the developed instrument has good psychometric properties, though several items require revision and refinement.

3.1. Content Validity

Based on the results of expert validation analysis from four experts, the results can be seen in Table 2.

Table 2. Content Validity of SPS Test Items

Component Science Process Skills	Average	Validity
Classifying	0,93	HS
Measurements	0,95	HS
Communication	0.90	HS
Observation	0.90	HS
Predicting	0,92	HS
Drawing conclusion	0,88	HS
Defining Operationally	0,88	HS
Identify variables	0,88	HS
Creating a hypothesis	0,9	HS
Data interpretation	0,88	HS
Changing and controlling variables	0,85	HS
Designing an experiment	0,88	HS
Total Average	0,9	HS

Expert judgments revealed that all 28 test items exhibited high content validity, with average ratings ranging from 0.85 to 0.95, resulting in an overall mean score of 0.90. These results confirm that the developed instrument adequately measures the intended science process skills (SPS) constructs. Minor revisions were made based on expert comments, particularly regarding the clarity of the wording and punctuation.

Importantly, the items also incorporate environmental contexts, such as analyzing the impact of salt-containing household wastewater on water pH. This contextual integration strengthens the instrument's ecological validity and aligns with science education priorities emphasizing socioscientific issues. Similar findings were reported by Subagja et al. (2023), who demonstrated that SPS test items based on real-life scientific phenomena had high content validity and contextual relevance.

3.2. Item Discrimination

The differentiability of the items is calculated based on the results of the trial on 150 students based on the upper and lower grades. The results of the item differentiability analysis can be seen in Table 3.

Table 3. Discrimination Index of SPS Items

Component Science Process Skills	Question No.	Skor	Differential Power
Classifying	1	0,47	Good
	2	0,64	Good
	3	0,27	Sufficient
Measurements	9	0.73	Very good
	10	0.56	Good
	24	0.62	Good
Communication	12	0.16	Poor
	21	0.67	Good
Observation	27	0.60	Good
	28	0.62	Good
Predicting	4	0.69	Good
	5	0.51	Good
	23	0.58	Good
Drawing conclusion	8	0.58	Good
	11	0.49	Good
	22	0.51	Good
Defining Operationally	19	0.51	Good
	20	0.69	Good
Identify variables	25	0.58	Good
	26	0.69	Good
Creating a hypothesis	6	0.76	Very good
	7	0.58	Good
Data interpretation	13	0.82	Very good
	14	0.87	Very good
Changing and controlling variables	15	0.51	Good
	16	0.24	Sufficient
Designing an experiment	17	0.49	Good
	18	0.71	Very good

The item discrimination analysis revealed that five items were classified as very good, 20 as good, two as sufficient, and one as poor. Item 12, which measured communication skills, was the lowest-performing item with a discrimination index of 0.16 (poor).

An example of a low-discrimination item is Item 12, which measures communication skills and obtained a discrimination index of 0.16.

"Describe how you would explain changes in water pH due to salt hydrolysis to a community group unfamiliar with chemistry concepts."

This item likely underperformed due to the dual demands of conceptual understanding and communication ability. Similar challenges were noted in Karim et al. (2024), where items requiring higher-order communication and interpretation produced lower discrimination indices despite strong overall psychometric results. It is implied that such items should be refined by providing scaffolds or structured response formats to improve their ability to differentiate between high- and low-achieving students.

3.3. Item Difficulty

The level of difficulty of the items is calculated based on the results of the trial on 150 students. The results of the analysis of the level of difficulty of the items can be seen in Table 4.

Table 4. Difficulty Index of SPS Items

Component Science Process Skills	Question No.	Skor	Differential Power
Classifying	1	0,84	Easy
	2	0,57	Medium
	3	0,35	Medium
Measurements	9	0.57	Medium
	10	0.41	Medium
	24	0.63	Medium
Communication	12	0.17	Difficult
	21	0.54	Medium
Observation	27	0.43	Medium
	28	0.61	Medium
Predicting	4	0.55	Medium
	5	0.57	Medium
	23	0.74	Easy
Drawing conclusion	8	0.49	Medium
	11	0.49	Medium
	22	0.47	Medium
Defining Operationally	19	0.47	Medium
	20	0.67	Medium
Identify variables	25	0.52	Medium
	26	0.62	Medium
Creating a hypothesis	6	0.58	Medium
	7	0.68	Medium
Data interpretation	13	0.65	Medium
	14	0.58	Medium
Changing and controlling variables	15	0.44	Medium
	16	0.31	Medium
Designing an experiment	17	0.69	Medium
	18	0.47	Medium

The results of the item difficulty analysis showed that two items were categorized as easy, 25 as medium, and one as difficult. The predominance of medium-difficulty items suggests that the test provides an appropriate level of challenge for students.

An example of a medium-difficulty item is Item 3, which tests classification skills.

"Classify the following salt solutions—NaCl, NH_4Cl, and Na_2CO_3—as acidic, neutral, or basic after hydrolysis. Provide explanations."

An example of a difficult item (Item 12 – Communication Skill) is:

"You measured a decrease in river water pH due to household salt discharge. Communicate your findings and explain the environmental implications to local residents."

Students found items that required synthesis of conceptual knowledge with real-world communication more difficult. This finding is consistent with Subagja et al. (2023), who noted that items contextualized in socioscientific issues tend to increase cognitive demand while enhancing assessment authenticity.

3.4. Item Validity

All 28 items achieved statistically significant validity, with correlations exceeding the critical threshold ($r = 0.159$, $N = 150$, $\alpha = 0.05$). One notable benchmark is the I-KPS instrument developed by Jalil et al. (2018). This instrument assessed six SPS indicators using essay items across various physics topics. The researchers reported a content validity coefficient of 0.96, and they empirically validated all 45 items, demonstrating strong fit and reliability ($\alpha = 0.935$). This demonstrates that instruments combining statistical validity with domain relevance, like yours, are both rigorous and meaningful.

3.5. Reliability Test

The reliability analysis yielded a Cronbach's alpha of 0.891, indicating very high internal consistency. This result shows that the developed test is a reliable tool for assessing SPS. Tosun (2019) reported similar reliability when developing an SPS test on the topic of "Matter and Its Nature," with Cronbach's alpha values of 0.80 for the total test and 0.64–0.71 for the subscales. Compared with these benchmarks, the reliability of the present instrument is notably strong.

3.6. Integration of Environmental Chemistry Concepts

In addition to psychometric quality, the instrument highlights the environmental chemistry concept of salt hydrolysis, which is closely related to real-world issues. One relevant problem is the change in water pH due to the discharge of household waste containing salts. For example, detergents (containing sodium carbonate, Na_2CO_3) can cause wastewater to become basic after hydrolysis, while fertilizers containing ammonium (NH_4^+) can lower water pH upon hydrolysis. These fluctuations directly affect aquatic ecosystems by influencing heavy metal solubility, disturbing ecological balance, and threatening aquatic life.

As an illustration, one example multiple-choice question developed in this study is as follows:

Example Question (Prediction and Data Interpretation – C-SPS and E-SPS):

After many households in a residential area discharge wastewater containing NH_4Cl, the pH of a nearby river decreases from 7 to 5. Which of the following best explains this phenomenon?

A. NH_4Cl hydrolyzes to form OH^- ions, increasing the pH of the water.

B. NH_4^+ ions hydrolyze to produce H_3O^+ ions, which lowers the pH of the water.

C. Cl^- ions react with water to form HCl, which increases the pH.

D. NH_4Cl does not hydrolyze in water, so the pH remains neutral.

Correct answer: B.

$NH_4Cl \rightarrow NH_4^+ + Cl^-$; $NH_4^+ + H_2O \rightleftharpoons NH_3 + H_3O^+$, which causes the solution to become acidic.

Environmental Impact: A decrease in pH (acidic water) increases the solubility of heavy metals (e.g., Fe, Pb, and Cd), which is harmful to aquatic organisms and human health. Students may also be asked to design a simple experiment using NH_4Cl solutions of varying concentrations to measure pH and make a comparison with a control group.

Relevance to SPS Instrument:

a. Observation and Measurement (B-SPS): Students conduct pH measurements on solution samples.
b. Identifying Variables (C-SPS): Students determine the independent variable (salt concentration), dependent variable (pH), and controlled variables (volume and temperature).
c. Hypothesis Making (E-SPS): Students formulate a hypothesis, such as, "The higher the NH_4Cl concentration, the lower the solution's pH."
d. Experimental Design (E-SPS): Students design simple steps to verify the phenomenon.
e. Data Interpretation (E-SPS): Students analyze the pH data and connect it to the hydrolysis reaction to explain the environmental implications.

Compared with similar SPS instrumeion, making it more relevant to real-world applications. For example, instruments developed by Subagja et al. (2023) that are grounded in socioscientific issues, such as cellular biology in community contexts, have demonstrated that incorporating real-life phenomena helps students engage more deeply and recognize the importance of science in their surroundings. Another study by Karim et al. (2024) utilizing Rasch modeling confirmed that SPS items based on everyday science phenomena yield excellent person and item reliability (0.97 and 0.99, respectively) with strong fit indices in MNSQ and infit metrics. These benchmarks demonstrate that the current instrument achieves strong psychometric properties and embeds authentic environmental contexts, such as salt hydrolysis impacting water pH, thereby strengthening cognitive and ecological validity.

From a pedagogical perspective, teachers can use this instrument to diagnose students' specific SPS weaknesses. For instance, if many students have difficulty with communication-oriented items, such as item 12, instruction could be reinforced with structured discussions, scientific presentations, or scaffolded writing tasks. This approach aligns with broader findings in science communication. For instance, Frontiers in Psychology analyzed how scientific literacy and communication around socio-scientific issues promote responsible global citizenship (Li & Guo, 2021) and Leuzinger et al. (2019) demonstrated that technology-mediated tools, such as smart device apps, effectively enhance climate change literacy and scientific communication. Thus, integrating real-world environmental issues into SPS assessments strengthens diagnostic precision and nurtures students' ability to communicate scientifically about pressing global topics.

4. Conclusion

The SPS test instrument on salt hydrolysis material developed showed a strong psychometric properties. First, the validity of the items based on expert judgment resulted in a high content validity index (CVI= 0.9), which means the questions are almost in perfect category. The trial results showed that 28 questions were valid, yielding a reliability coefficient of 0.891 which means high internal consistency and reliability. The index of difficulty of the question instrument showed that 2 questions fell into the easy category, 25 questions into the medium category and 1 question into the difficult category.

These results indicate that this instrument is suitable for various levels of student ability and aligns well with the SPS indicators. The test questions can be used to measure science process skills and for formative and summative evaluations. Teachers can use this instrument to make valid and reliable assessments. This research can serve as a reference for developing similar instruments in other subjects. This study only focused on the hydrolysis of salts in the 11th grade chemistry curriculum. The developed instruments were limited to multiple-choice questions and did not include performance-based assessments or hands-on laboratory experiments. Further testing of the instruments developed in various schools with different student characteristics is needed to determine the consistency and generalizability of the test results. Subsequent research could analyze the effect of using these science process skills assessment instruments on student learning outcomes, including cognitive and affective outcomes.

Author Contributions

All authors have equal contributions to the paper. All the authors have read and approved the final manuscript.

Funding

No funding support was received.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). 'An experiment is when you try it and see if it works': A study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, *11*(5), 514–529. https://doi.org/10.1080/0950069890110504

Chang, R., & Goldsby, K. A. (2016). *Chemistry* (12th editi). McGraw-Hill Education.

Crandell, O. M., et al. (2024). Open-ended Green Chemistry Generic Misconception Prompt: Instrument development and validation. *Journal of Chemical Education*, *101*(5), 1204–1215.

Damanhuri, M. I. M., Treagust, D. F., Won, M., & Chandrasegaran, A. L. (2016). High school students' understanding of acid-base concepts: An ongoing challenge for teachers. *International Journal of Environmental and Science Education*, *11*(1), 9–27. https://doi.org/10.12973/ijese.2015.284a

Danczak, S. M., Thompson, C. D., & Overton, T. L. (2020). Development and validation of an instrument to measure undergraduate chemistry students' critical thinking skills. *Chemistry Education Research and Practice*, *21*(1), 62–78. https://doi.org/10.1039/c8rp00130h

Deng, J. M., Streja, N., & Flynn, A. B. (2021). Response Process Validity Evidence in Chemistry Education Research. *Journal of Chemical Education*, *98*(12), 3656–3666. https://doi.org/10.1021/acs.jchemed.1c00749

Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics - Physics Education Research*, *5*(2). https://doi.org/10.1103/PhysRevSTPER.5.020103

Ding, L., Wei, X., & Liu, X. (2022). Scientific reasoning as a predictor of inquiry learning outcomes. *Journal of Research in Science Teaching*, *59*(3), 402–426.

Godec, S., & King, H. (2021). Developing a science engagement scale using Rasch analysis. *International Journal of Science Education*, *43*(18), 3012–3035.

González-Cabrera, J., Brunner, M., & De Boeck, P. (2021). A theoretical model and quantitative assessment of scientific thinking. *Physical Review Physics Education Research*, *18*(1), 010115.

Grieger, A. J., Graulich, N., & Schmid, S. (2022). Assessing students' knowledge of green chemistry principles: Development and validation of the ASK-GCP inventory. *Chemistry Education Research and Practice*, *23*(4), 1226–1245.

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

Jalil, S., Herman, Sidin Ali, M., & Haris, A. (2018). Development and validation of science process skills instrument in physics. *Journal of Physics: Conference Series*, *1028*(1). https://doi.org/10.1088/1742-6596/1028/1/012203

Karim, E., Ibrahim, S. I., Rasit, H. H., & Sumintono, B. (2024). *Development and Validation of Students' Competency Instrument on Science Process Skills* (Issue Proms 2023). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-494-5_20

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, *65*(23), 2276–2284. https://doi.org/10.2146/ajhp070364

Landis, J. R., & Koch, G. G. (1977). Landis amd Koch1977_agreement of categorical data. *Biometrics*, *33*(1), 159–174.

Lazonder, A. W., & Janssen, N. (2021). Development and initial validation of a performance-based scientific reasoning test for children. *Studies in Educational Evaluation*, *68*(December 2020), 100951. https://doi.org/10.1016/j.stueduc.2020.100951

Leuzinger, S., Borrelle, S. B., & Jarvis, R. M. (2019). Improving Climate-Change Literacy and Science Communication Through Smart Device Apps. *Frontiers in Education*, *4*(November), 1–5. https://doi.org/10.3389/feduc.2019.00138

Lewis, R., & Evans, W. (2018). *Chemistry* (5th editio). Red Globe Press.

Li, Y., & Guo, M. (2021). Scientific Literacy in Communicating Science and Socio-Scientific Issues: Prospects and Challenges. *Frontiers in Psychology*, *12*(November). https://doi.org/10.3389/fpsyg.2021.758000

Liu, S., Xu, S., Li, Q., Xiao, H., & Zhou, S. (2023). Development and validation of an instrument to assess students' science, technology, engineering, and mathematics identity. *Physical Review Physics Education Research*, *19*(1), 10138. https://doi.org/10.1103/PhysRevPhysEducRes.19.010138

McMurry, J. E., Fay, R. C., & Robinson, J. K. (2015). *Chemistry* (7th Editio). Pearson.

Mede, N. G., Howell, E. L., Schäfer, M. S., Metag, J., Beets, B., & Brossard, D. (2025). Measuring Science Literacy in a Digital World: Development and Validation of a Multi-Dimensional Survey Scale. *Science Communication*, *X*. https://doi.org/10.1177/10755470251317379

Mi, S., Ye, J., Yan, L., & Bi, H. (2023). Development and validation of a conceptual survey instrument to evaluate senior high school students' understanding of electrostatics. *Physical Review Physics Education Research*, *19*(1), 10114. https://doi.org/10.1103/PhysRevPhysEducRes.19.010114

MoEC. (2016). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 24 Tahun 2016 tentang Standar Isi Pendidikan Dasar dan Menengah*. Kementerian Pendidikan dan Kebudayaan Republik Indonesia.

Muntholib, Ibnu, S., Rahayu, S., Fajaroh, F., Kusairi, S., & Kuswandi, B. (2020)a. Chemical literacy: Performance of first year chemistry students on chemical kinetics. *Indonesian Journal of Chemistry*, *20*(2), 468–482. https://doi.org/10.22146/ijc.43651

Muntholib, Khusmawardani, E., Utomo, Y., Muchson, & Yahmin. (2020)b. Development and implementation of multiple-choice chemical literacy survey in acid-base chemistry. *AIP Conference Proceedings*, *2215*(April). https://doi.org/10.1063/5.0000547

Muntholib, Mauliya, A. H., Utomo, Y., & Ibnu, M. S. (2020)c. Assessing high school student's chemical literacy on salt hydrolysis. *IOP Conference Series: Earth and Environmental Science*, *456*(1). https://doi.org/10.1088/1755-1315/456/1/012065

Park, M., & Song, J. (2021). Development of an assessment tool for integrated science process skills using Rasch modeling. *Research in Science Education*, *51*(3), 763–783.

Pluye, P., et all. (2021). Ethical challenges in educational research with children. *Accountability in Research*, *28*(5), 291–307.

Regier, D. A., Kupfer, D. J., Kuhl, E. A., Ph, D., Greiner, L., Kraemer, H. C., & Ph, D. (2013). DSM-5 Field Trials in the United States and Canada, Part I: Study Design, Sampling Strategy, Implementation, and Analytic Approaches. *American Journal of Psychiatry*, *170(1)*, 59–70.

Semiawan, C., Tangyong, A., Belen, S., Matahelemudal, Y., & Suseloardjo, W. (1992). *Pendekatan Keterampilan Proses*. Gramedia Widiasarana Indonesia.

Subagja, S., Rubini, B., & Kurniasih, S. (2023). Development and Validation of a Test of Science Process Skills for Secondary Students on Cellular Living System Organization Matter: Rasch Model Analysis. *Jurnal Penelitian Pendidikan IPA*, *9*(10), 9056–9062. https://doi.org/10.29303/jppipa.v9i10.5177

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tosun, C. (2019). Scientific process skills test development within the topic "Matter and its Nature" and the predictive effect of different variables on 7th and 8th grade students' scientific process skill levels. *Chemistry Education Research and Practice*, *20*(1), 160–174.

van Griethuijsen, R. A. L. F., van Eijck, M. W., Haste, H., den Brok, P. J., Skinner, N. C., Mansour, N., Gencer, A. S., & BouJaoude, S. (2015). Global patterns in students' views of science and interest in science. *Research in Science Education*, *45*(4), 581–603. https://doi.org/10.1007/s11165-014-9438-6

Wuttiprom, S., Sharma, M. D., Johnston, I. D., Chitaree, R., & Soankwan, C. (2009). Development and use of a conceptual survey in introductory quantum physics. *International Journal of Science Education*, *31*(5), 631–654. https://doi.org/10.1080/09500690701747226

Yan, Z., & Pastore, R. (2022). Validation practices in educational measurement: A systematic review. *Journal of Psychoeducational Assessment*, *40*(7), 866–884.

Ye, L., Lewis, M. M., & Chen, X. (2023). Measuring STEM identity: Development and validation of a multidimensional scale. *Hysical Review Physics Education Research*, *19*(1), 010138.

Zhang, Y., Xu, X., & Lin, T. (2020). Assessing inquiry-based science learning outcomes using Rasch analysis. *Research in Science Education*, *50*(5), 2021–2038.

Zhou, S., & Lee, M. K. (2021). Development and validation of a scientific modeling competency test. *International Journal of Science Education*, *43*(15), 2461–2482.