



Development and Implementation of Students' Scientific Argumentation Skills Test in Acid-Base Chemistry

Parlan*, Umi Latifah, Muntholib

Faculty of Mathematics and Sciences, Universitas Negeri Malang, Indonesia

Article History:

Received: May 13th, 2020

Revised: July 27th, 2020

Accepted: August 19th, 2020

Published: December 31st, 2020

Keywords:

Acid-base chemistry,
Conceptual understanding,
High school students,
Scientific argumentation

*Correspondence Address:

parlan.fmipa@um.ac.id

Abstract: This study aimed to develop and validate Scientific Argumentation Skills Test (SAST) and investigate the 11th-grade students' performance in scientific argumentation skills on acid-base chemistry. The research design used was research and development, followed by descriptive research. Research and development were carried out to obtain an instrument of SAST, and descriptive research was used to describe students' argumentation skills in acid-base chemistry. Participants in this study were 328 11th-grade students of state high schools in East Java, Indonesia. The research and development of SAST consisted of five steps, namely literature review, items development, expert judgment, pilot project, and finalization of instruments. Expert judgment involved three chemistry education experts, while the pilot project involved 151 students, and the identification of students' scientific argumentation skills involved 177 students. Data about expert assessments, student responses to the pilot project, and student answers to the application of SAST were analyzed descriptively. The SAST produced in the research and development steps consisted of parts A (10 items) and part B (7 items), with Cronbach's alpha reliability coefficients of 0.888 and 0.758. The students' performance in scientific argumentation skills showed that the average score of students' performance to determine the argument's components was 80.53 % (excellent category). The average score of students' performance to write an argument was 55.42 % (moderate category). The implication of the study that the students' scientific argumentation skills must be explicitly trained in learning.

INTRODUCTION

Argumentation skills are an essential aspect of learning. Students' argumentation skills relate to their understanding. The development of argumentation skills may influence conceptual understanding (Albe, 2008; Choden & Kijkuakul, 2020; Harris & Ratcliffe, 2005). Learning that facilitates students to improve their argumentation skills will increase their understanding of the material being studied. The research findings also revealed that the knowledge

level is influential in argumentation skills (Demiral & Cepni, 2018). Therefore, the use of innovative learning models that are appropriate to the characteristics of the learning material can eventually empower students' argumentation skills (Noviyanti et al., 2019). The results of the study showed that the academic ability of students influences the way students construct arguments. Students who have high academic abilities have better argumentation skills than students who have low academic abilities because high-

ability students are more skilled in gathering data and evidence (Nurramadhani, 2017). The learning that facilitates the development of academic abilities also increases the ability to argue. Students with high academic abilities can also communicate the results (Kollar et al., 2014; Osborne et al., 2004).

Argumentation skill is one factor that determines student success in school (Frey et al., 2015). Argumentation is a component of scientific knowledge, which aims to learn, learn, and construct scientific knowledge (Erduran, 2007). The ability to integrate knowledge and ideas, describe and evaluate claims and arguments, and assess the reasons used in arguments is central to the Common Core State Standards. The process of argumentation is a central component of science education that will help students make decisions now and in the future. In Indonesia, the students' ability to argue is developed in learning through a scientific approach (Competency-based Curriculum 2013). The consequence of placing scientific arguments as objectives and ways of learning science is the availability of instruments for evaluating scientific arguments (Sampson & Blanchard, 2012). The results showed that linking learning practices to encourage students to argue could improve students' argumentation abilities (Grooms et al., 2010).

Argumentation is an essential aspect of learning science (chemistry). There is a link between the ability to argue with academic achievement. Scientific arguments can help increase students' knowledge about concepts, involvement in scientific work, and literacy. To be able to argue well requires mastery of scientific inquiry and scientific literacy. A fundamental variable in scientific inquiry and scientific literacy are integrated science process skills (Rauf et al., 2013). The importance of argumentation in academic achievement was recognized by many researchers in the literature (Kosko et al., 2014; Walter & Barros, 2011). The

study results showed that students have difficulty constructing scientific arguments to improve their understanding of knowledge (Heng et al., 2014). The students also have some problems in mastering scientific argumentation (Choden & Kijkuakul, 2020). Therefore, efforts to link the ability to argue with the scientific knowledge of science need to be made.

Understanding chemical concepts and the interrelationships between concepts will enhance scientific argumentation skills. Chemistry is a part of science, having many interrelated to students' argumentation skills and understanding (Walter & Barros, 2011). Learning of science facilitates students to construct scientific assumptions about phenomena accompanied by appropriate evidence and scientific principles (McNeill & Krajcik, 2008). Argumentation indicators are needed to assess learning. Tests with argumentation patterns can train students' argumentation skills and measure students' understanding of the material being studied to be used as a reference for improving the learning process and chemical assessment (Osborne et al., 2004).

From several decades ago until now, the acids and bases topic has been reported to be difficult for high school students who have, as a result, held several alternative conceptions about acids and bases (Damanhuri et al., 2016). In the study of acid-base chemistry, students' understanding is built through observation (data collection) to explain phenomena related to the results of observations. Students' understanding is closely related to the ability to explain phenomena based on observational results. An understanding of acid-base relates to scientific argumentation components consisting of claims, evidence (observational and measurement data), and explanations (how evidence can support claims). Science learning

facilitates students to build scientific arguments about phenomena equipped with appropriate evidence and scientific principles (McNeill & Krajcik, 2008). Scientific explanations answer three questions, namely what is known (ontology questions), why they occur (causal questions), and how to know (epistemic questions) (Osborne et al., 2004).

Braaten & Windschitl (2011) developed a guide that teachers can use to assess the depth of scientific argumentation made by students. Furtak et al. (2010) developed a framework used to analyze scientific arguments/explanations by adopting the Toulmin framework (Osborne et al., 2004). Scientific arguments involve complex relationships among many skills (Brown et al., 2010). Brown et al. (2010) developed a method for measuring students' ability to provide scientific explanations.

Scientific argumentation skills assessments have been developed by Frey et al. (2015) and Grooms et al. (2010). Scientific argumentation skills test developed only to measure one component of argumentation skills, namely constructing arguments containing claims, evidence, and explanations. However, the tests have not been equipped with an assessment of students' necessary skills to distinguish

the components of argumentation. Until now, there has not been developed an argumentation skills test that measures the two components of argumentation skills, namely the students' basic skills in distinguishing argumentation components and the students' skills of writing scientific arguments, especially in acid-base chemistry

Necessary aim of this study were to study the 11th-grade students' argumentation skills in acid-base chemistry. These objectives are broken down into two stages: 1) developing and validating the argumentation skills test (SAST) on acid-base chemistry, 2) investigating the performance of the 11th-grade students in scientific argumentation skills on acid-base chemistry.

METHOD

Research Design

The study used two types of research designs: 1) research and development design (R and D) to produce the Scientific Argumentation Skills Test (SAST) on acid-base chemistry, 2) descriptive research design to describe the performance of high school students in scientific argumentation skills on acid-base chemistry. The SAST development steps to adapt the methods carried out by (Wattanakasiwich et al., 2013) and (Chandrasegaran et al., 2007) with some changes, as shown in Figure 1.

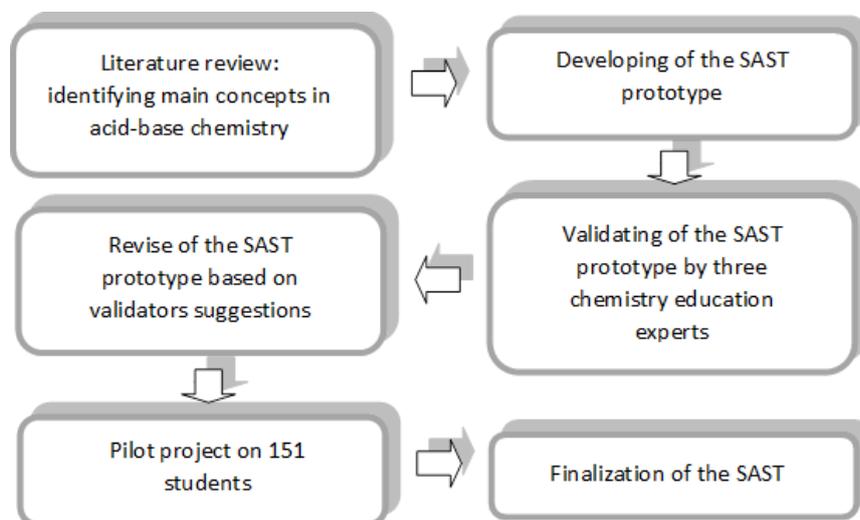


Figure 1. The Development Steps of the SAST

The purpose of stage 1 was to identify the acid-base chemistry concepts used to construct argumentation questions in the study. These include the history of acid-base concepts, acid-base indicators, and pH calculations to produce concept maps of acid-base material. The literature review aimed to define the scope of content that will be explored of the questions that should be provided.

The purpose of stage 2 is to construct a prototype of SAST and its

scoring rubric. The SAST consists of two parts. Part A contains the statements in which students are asked to classify each statement into claims (conclusions), evidence (supporting conclusions), or explanations. Part B contains a description where students are asked to make claims, evidence, and explanations based on the description. Examples of questions for parts A and B are shown in Figure 2.

Problem Part A:
The ionization constant value of water is affected by changes in temperature. (Is the statement a claim, evidence, or explanation?)

Problem Part B.
Range of pH some indicators are as follows.

Indicator	Range of pH	Discoloration	
		Acid	Base
Methyl orange (MO)	3.1-4.4	Red	Yellow
Methyl red (MM)	4.4-6.2	Red	Yellow
Bromothymol blue (BTB)	6.0-7.6	Yellow	Blue
Phenolphthalein (PP)	8.3-10.0	Colorless	Red purple

A student conducts an experiment to test solution A whose pH is unknown. Following are the results of the student's experiments on solution A.

Added indicator	Color
Methyl red (MR)	Orange
Methyl orange (MO)	Yellow
Bromothymol blue (BTB)	Yellow
Phenolphthalein (PP)	Colorless

a) Based on these data, what is the pH of solution A?
.....

b) What evidence do you use to support your answer?
.....

c) Explain why the evidence supports your answer?
.....

Figure 2. Examples of Questions for Parts A and B

The purpose of stage 3 is to validate the SAST by three chemistry education experts. In validation, the three chemistry education experts assess each item of the SAST and provide suggestions for improvement if needed. In stage 4, a pilot project is carried out to obtain information about the validity, different power, and

the difficulty level of items and information about its reliability. The pilot project was conducted on 151 11th-grade senior high school students. After data collection, the instrument was analyzed in terms of the Difficulty Level (DL), Discriminatory Index (DI), validity, and reliability. The criteria used to interpret

the items analysis parameter is shown in Table 1. Items validity and reliability are determined by the SPSS program using Product Moment correlation and Cronbach alpha (Kimberlin & Winterstein, 2008), respectively. The purpose of stage 5 was to finalize the SAST based on pilot project data.

Participants

Participants of the study were 328 11th-grade students of state high schools in Malang, East Java, Indonesia, with details, 151 students for the instrument's pilot project, and 177 students for the

identification of argumentation skills. Participants' selection was carried out using purposive sampling techniques, namely 11th-grade students who had obtained acid-base material.

Data Collection and Analysis

Three chemistry education experts carried out the SAST validation. The assessment is carried out on five aspects: content, the rubric of assessment, construct, display and applicability, and language. The assessment is carried out with a Likert scale of 1-4 with the criteria, as shown in Table 2.

Table 1. The Criteria Used to Interpret the Item Analysis Parameter

Validity		Discrimination Index (DI)		Difficulty Level (DL)		Reliability	
	criteria	value	criteria	value	criteria	value	criteria
$R_{xy-tab}=0.220$	valid	0.00 - < 0.10	poor	0.00 - 0.30	Hard	> 0.90	excellent
$R_{xy} > r_{xy-tab}$	invalid	0.10 - < 0.30	fair	0.31 - 0.70	Moderate	0.81 - 0.90	very good
$R_{xy} < r_{xy-tab}$		0.30 - < 0.75	good	0.71 - 1.00	Easy	0.61 - 0.80	good
		0.75 - 1.00	excellent			0.40 - 0.60	fair
		negative	unsuitable			< 0.4	poor

Table 2. Criteria for Assessment of the SAST

Score	Category
1	Item is irrelevant or inappropriate
2	Item items are less relevant or inappropriate
3	Item is relevant or appropriate
4	Item is very relevant or very appropriate

The score given by each validator to each indicator is expressed in percentages by the formula:

$$\% \text{ score} = \frac{\sum \text{total score obtained}}{\sum \text{maximum score}} \times 100 \%$$

The feasibility level of the developed instrument is identified based on the percentage of content validity of each item calculated by the formula:

$$\text{feasibility level} = \frac{\text{average } \sum \text{total score per item}}{\sum \text{maximum score per item}} \times 100 \%$$

Based on the percentage of content validity, each item's validity can be determined by criteria.

Table 3. Criteria for Percentage of Content Validity

Percentage (%)	Category
80.1 - 100	Very Worthy
60.1 - 80	Feasible
40.1 - 60	Fair
20.1 - 40	Not Eligible
0 - 20	Very Ineligible

(Arikunto, 2009)

The pilot project on 155 students conducted empirical validity, discrimination index, the difficulty level of the item, and reliability of the SAST. Analysis of validity of item and reliability was carried out by SPSS program for Windows version 23.0 ($p < 0.05$). Discrimination index and difficulty level of items were carried out manually by the Excel program.

The SAST prototype that has fulfilled the criteria of validity, discrimination index, difficulty level, and reliability is used to identify high school students' argumentation skills. Participants in this activity were 177 11th-grade students of two public schools in

Malang, Indonesia. Both schools use the same curriculum and textbooks. Students' answers are given a score to calculate the percentage of each question. The scoring of students' answers is done by two persons independently. The criteria used to determination of mastery level of students' scientific argumentation skills are presented in Table 4 (Heng et al., 2014).

Table 4. The Criteria Used to Determination of Mastery Level

Mean Score (%)	Mastery Level
80.00 - 100.00	Excellent
60.00 - 79.99	Good
40.00 - 59.99	Moderate
20.00 - 39.99	Weak
00.00 - 19.99	Very Weak

RESULT AND DISCUSSION

Prototype of SAST

The product of the research and development is in the form of the

Scientific Argumentation Skills Test (SAST) prototype on acid-base chemistry. The items in the SAST were developed based on the assessment aspects recommended by Frey et al. (2015). The SAST structure is also adjusted to the argument framework, according to (Sampson & Blanchard, 2012; Sampson & Clark, 2008; Schleigh, 2014). The SAST prototype consisted of two parts, namely Part A (15 items) and Part B (10 items). Problem Part A measures the students' skills to identify whether a statement includes claims, data, or reasoning/explanation. Problem Part B measures students' ability to construct scientific explanations that contain components of claim, data, and explanation. The SAST prototype validation results by the chemistry education experts for each indicator are presented in Table 5.

Table 5. Percentage of Feasibility of SAST Prototype for Each Indicator/Aspect

Indicator/Aspect	Percentage of Score		Average	Category
	Part A	Part B		
Content	81.67	91.67	86.67	Very worthy
Rubric of assessment	87.50	87.50	87.50	Very worthy
Construct	88.89	86.11	87.50	Very worthy
Display and applicability	80.56	88.89	84.72	Very worthy
Language	77.78	77.78	77.78	Feasible
Average	83.28	86.39	84.83	Very worthy

Based on Table 5, it is known that the SAST prototype obtained an average percentage of 84.83 % within the very feasible category. Criteria to be used as an assessment test for students' scientific argumentation skills. Validators were also given some suggestions to improve/revise the SAST. The validators suggestions are *problem section B No.7, the question is clarified by changing one of the observational evidence; problem section B No.10, the value of the log of 3.44 needs to be included in the problem in order to facilitate student calculations; problem part A No. 1, use of the word "must" be omitted; problem Part A No. 14, "the degree of ionization is not a constant";*

and an example of working on part B should be given.

Items Analysis (The Result of Pilot Project of the SAST Prototype)

Analysis of items includes validity, discrimination index, difficulty level, and reliability. The validity, discrimination index, and difficulty level of SAST Part A and Part B are presented in Table 6, Table 7, and Table 8.

Analysis of validity showed that there were three invalid questions and 12 valid questions for Part A and all valid questions for Part B. Discriminatory Index (DI) analysis showed that there were two low questions, two fair questions, 11 useful questions for Part A,

and Part B there was one flawed question, three fair questions, four right questions, two excellent questions. Difficulty Level (DL) analysis showed that there were 13 easy questions, two moderate questions for Part A and three easy questions, six moderate questions, one hard question for Part B. Reliability of the SAST Prototype is presented in Table 9.

Students' Performance in Scientific Argumentation Skills on Acid-Base Chemistry

Students' performance in scientific argumentation skills on acid-base chemistry is shown in two aspects of performance: the ability to distinguish the components of argumentation skills (claim, evidence, explanation) and construct scientific explanations. The performance of students' scientific argumentation skills to distinguish each argumentation component is shown in Table 10. Students' performance in constructing scientific arguments is presented in Figure 3.

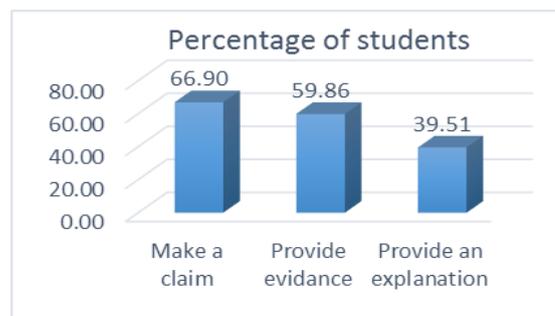


Figure 3. Students' Performance in Constructing Scientific Arguments

The percentage of students who can construct scientific arguments related to acid-base chemistry concepts is shown in Figure 3. Based on the results of the validity, DI, and DL analysis, ten questions of Part A (number 3, 4, 5, 6, 8, 9, 11, 12, 13, and 14) and seven questions of Part B (number 3, 4, 5, 7, 8, 9, and 10) were selected.

The reliability of Part A and Part B questions are shown in Table 9. Based on the SAST reliability listed in Table 9, it is known that the SAST is appropriate to be used to measure students' argumentation skills on acid-base chemistry.

Table 6. The Validity of Each Item/Question of the SAST Prototype with a Confidence Level of 95 % ($r_{xy-tab} = 0.159$)

Question	1	2	3	4	5	6	7	8	9	10
Part A	0.034 invalid	0.125 invalid	0.5798 valid	0.436 valid	0.452 valid	0.640 valid	0.144 valid	0.304 valid	0.400 valid	0.048 invalid
Part B	0.254 valid	0.406 valid	0.583 valid	0.534 valid	0.607 valid	0.339 valid	0.410 valid	0.640 valid	0.711 valid	0.762 valid

Table 7. The Discrimination Index (DI) of Each Item/Question of the SAST Prototype

Question	1	2	3	4	5	6	7	8	9	10
Part A	0.02 poor	0.38 good	0.35 good	0.60 good	0.42 good	0.40 good	0.19 fair	0.60 good	0.67 good	0.06 poor
Part B	0.09 poor	0.14 fair	0.36 good	0.31 good	0.41 good	0.15 fair	0.28 fair	0.49 good	0.79 excellent	0.80 excellent

Table 8. The Difficulty Level (DL) of Each Item/Question of the SAST Prototype

Question	1	2	3	4	5	6	7	8	9	10
Part A	0.99 easy	0.66 moderate	0.86 easy	0.79 easy	0.84 easy	0.88 easy	0.90 easy	0.68 moderate	0.76 easy	0.92 easy
Part B	0.62 moderate	0.92 easy	0.72 easy	0.64 moderate	0.24 hard	0.70 easy	0.58 moderate	0.42 moderate	0.60 moderate	0.33 moderate

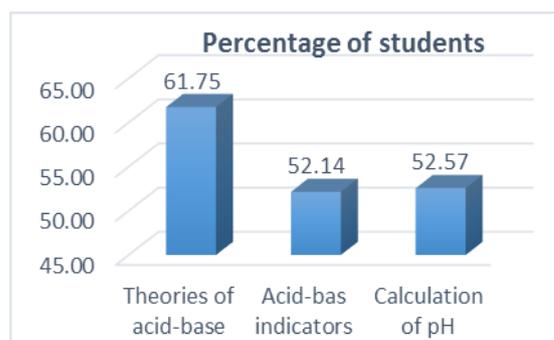
Table 9. Reliability of the SAST Prototype Using Cronbach Alpha with Confidence Level 95 %

Part	Reliability	Category
A (10 items)	0.888	Very good
B (7 items)	0.758	Good

Table 10. Performance of Students' Scientific Argumentation Skills to Distinguish Each Argumentation Component

Indicator	Percentage of Students			
	Claim	Evidence	Explanation	Average
Students can determine the statements included in claims, evidence, and explanations.	85.71	71.07	84.29	80.53
Category	Very good	Good	Very good	Very good

Based on Table 10, it is known that most students were able to distinguish statements, including claims, evidence, or explanations. Students could recognize statements that include claims, evidence, or explanations. Giving examples in the instructions manual of the test part A that is equipped with a description of each component of scientific argumentation helps students understand scientific argumentation to explain why the evidence and or claim support the scientific argumentation. The skill to distinguish each argumentation component can help students write scientific argumentations (Sampson & Gleim, 2009).

**Figure 4.** Percentage of Students Arguments Related to Concepts in Acid-Base Chemistry

Based on Figure 3, it is known that students' skills in making claims are better than deciphering evidence and making explanations. Students' skills in giving claims are better than two other argumentation skills. To make claims, students did not need to think critically, but students only need to recognize, remember, understand, classify, and re-apply their knowledge. In deciphering evidence, students are required to analyze and describe information that includes

evidence and not evidence. In contrast, in taking up an explanation, students are required to provide explanations that connect claims and evidence (why evidence can support claims) with scientific principles. Students' skills in elaborating evidence are also higher than in providing explanations. Student performance to explain is lowest compared to two other argumentation skills aspects. Most students have difficulty in delivering relevant explanations for why evidence can support claims. The teacher must facilitate students to understand essential concepts and the relationship between concepts to construct a good explanation. This is in line with previous studies' results that most students cannot provide scientific explanations well (Osborne et al., 2004). There was an improvement in the quality of students' argumentation.

Based on Figure 4, it is known that the students' argumentation skills in the sub-material the acidity of the solution (pH) are strong acids, strong bases, weak acids, and weak bases, and the acid-base indicator sub-materials are classified as sufficient. In contrast, the sub-material theories of acid-base are quite good. To understand the sub-material indicators of acid-base, students must understand the concepts of protonation and deprotonation. Most students can determine the acid-base indicator, but they cannot give an exact explanation related to the species that has discolored. They could not relate the phenomenon of color change to the acid-base reaction on the indicator. The same thing happens with sub material pH of strong acids, strong bases, weak acids, and weak bases.

Students must understand and use prior knowledge to classify weak acids, strong acids, weak bases, and strong bases. If this knowledge is not yet understood, students will have difficulty calculating the pH of acidic or basic solutions. Students' answers tend to give correct answers in calculating pH, but students have difficulty when they are asked to calculate other variables.

Acid-base chemistry is an essential element of the chemistry curriculum (Kooser et al., 2001). From as early as several decades ago, the topic of acids and bases has been difficult for high school students. Sampson et al. (2009) suggest using inquiry-based learning (e.g., Argument-Driven Inquiry/ADI) in science learning to improve understanding of concepts in science.

The leading cause of students' scientific argumentation skills is because they were less trained and accustomed to conveying their arguments. This is possible because teaching and learning activities and assessment used have not yet facilitated argumentation skills. The learning done by the teacher in the classroom is dominated by conventional learning models or expository, which only prioritizes the completeness of the material. Consequently, students lose the opportunity to argue in scientific discussions with peers (Paul & Elder, 2014). This causes the understanding of concepts that students have are incomplete. Therefore, teachers must use innovative learning models, such as inquiry-based learning, to overcome these weaknesses. Also, assessment tools are in the form of multiple-choice questions, short answers, or questions that prioritize algorithmic aspects. This causes students to memorize formulas so that students have difficulty explaining phenomena related to acid-base material (Cooper et al., 2016).

One of the goals of science education is to provide students with the ability to construct arguments-reasoning

and thinking critically in a scientific context (Katchevich et al., 2013). Arguments allow students to engage in various scientific practices in daily life through exploration activities during learning and increase their understanding of science's meaning (Tsai et al., 2015). The teachers need to choose a learning model that matches the material's characteristics that can facilitate students to construct scientific arguments. The teachers can also use social media (i.e., Facebook) to enhance their students' argumentation skills (Delen, 2017). Building an argument has significant social importance for students and their learning of scientific concepts (Katchevich et al., 2013). Students will gain scientific experience and its application so that it can be used to justify and support their arguments (Chowning et al., 2012).

CONCLUSION

The results of the study obtained from the SAST prototype are valid and reliable. The SAST consists of parts A (10 items) and part B (7 items), with the reliability of 0.888 and 0.758. The performance of students to distinguish statements including claims, evidence, or explanations were very good. Students' performance in constructing scientific argumentations was in a moderate category (55.42 %). Student performance to provide the explanation was the lowest compared to two other argumentation skills aspects. Most students have difficulty in delivering relevant explanations why evidence can support claims. The teachers need to practice the students' scientific argumentation skills explicitly in learning.

REFERENCES

Albe, V. (2008). When scientific knowledge, daily life experience, epistemological and social considerations intersect: Students' argumentation in group discussions

- on a socio-scientific issue. *Research in Science Education*, 38, 67–90. <https://doi.org/10.1007/s11165-007-9040-2>
- Arikunto, S. (2009). *Dasar-dasar evaluasi pendidikan*. Bumi Aksara.
- Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education*, 95(4), 639–669. <https://doi.org/10.1002/sce.20449>
- Brown, N. J. S., Nagashima, S. O., Fu, A., Timms, M., & Wilson, M. (2010). A framework for analyzing scientific reasoning in assessments. *Educational Assessment*, 15(3), 142–174. <https://doi.org/10.1080/10627197.2010.530562>
- Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, 8(3), 293–307.
- Choden, T., & Kijkuakul, S. (2020). Blending problem based learning with scientific argumentation to enhance students' understanding of basic genetics. *International Journal of Instruction*, 13(1), 445–462.
- Chowning, J. T., Griswold, J. C., Kovarik, D. N., & Collins, L. J. (2012). Fostering critical thinking, reasoning, and argumentation skills through bioethics education. *PLoS ONE*, 7(5), 1–8. <https://doi.org/10.1371/journal.pone.0036791>
- Cooper, M. M., Kouyoumdjian, H., & Underwood, S. M. (2016). Investigating students' reasoning about acid-base reactions. *Journal of Chemical Education*, 93(10), 1703–1712. <https://doi.org/10.1021/acs.jchemed.6b00417>
- Damanhuri, M. I. M., Treagust, D. F., Won, M., & Chandrasegaran, A. . (2016). High school students' understanding of acid-base concepts : An ongoing challenge for teachers. *International Journal of Environmental & Science Education*, 11(1), 9–27. <https://doi.org/10.12973/ijese.2015.284a>
- Delen, İ. (2017). Teaching argumentation by using facebook groups. *International Journal of Instruction*, 10(1), 151–168. <https://doi.org/10.12973/iji.2017.10110a>
- Demiral, Ü., & Cepni, S. (2018). Examining argumentation skills of preservice science teachers in terms of their critical thinking and content knowledge levels : An example using GMOs. *Journal of Turkish Science Education*, 15(3), 128–151. <https://doi.org/10.12973/tused.10241a>
- Erduran, S. (2007). Methodological foundations in the study of argumentation in science classrooms. In *Argumentation in Science Education*. Springer.
- Frey, B. B., Ellis, J. D., Bulgren, J. A., Hare, J. C., & Ault, M. (2015). Development of a test of scientific argumentation. *Electronic Journal for Research in Science & Mathematics Education*, 19(4), 1–18.
- Furtak, E. M., Hardy, I., Beimbrech, C., Shavelson, R. J., & Shemwell, J. T. (2010). A framework for analyzing evidence-based reasoning in science classroom discourse. *Educational Assessment*, 15(3), 175–196. <https://doi.org/10.1080/10627197.2010.530553>
- Grooms, J., Walker, J. O. I. P., & Al, S. E. T. (2010). Argument-driven inquiry as a way to help students learn how to participate in scientific

- argumentation and craft written arguments: An exploratory study. *Science Education*, 95(2), 217–257. <https://doi.org/10.1002/sce.20421>
- Harris, R., & Ratcliffe, M. (2005). Socioscientific issues and the quality of exploratory talk — what can be learned from schools involved in a ‘collapsed day’ project? *The Curriculum Journal*, 16(4), 439–453. <https://doi.org/10.1080/09585170500384396>
- Heng, L. L., Surif, J., & Seng, C. H. (2014). Individual versus group argumentation: Student’s performance in a malaysian context. *International Education Studies*, 7(7), 109–124.
- Katchevich, D., Hofstein, A., & Mamlok-naaman, R. (2013). Argumentation in the chemistry laboratory: Inquiry and confirmatory experiments. *Research in Science Education*, 317–345. <https://doi.org/10.1007/s11165-011-9267-9>
- 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>
- Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction*, 32, 22–36. <https://doi.org/10.1016/j.learninstruc.2014.01.003>
- Kooser, A. S., Jenkins, J. L., & Welch, L. E. (2001). Acid-base indicators: A new look at an old topic. *Journal of Chemical Education*, 78(11), 11–13.
- Kosko, K. W., Rougee, A., & Herbst, P. (2014). What actions do teachers envision when asked to facilitate mathematical argumentation in the classroom? *Math Ed Res J*, 26, 459–476. <https://doi.org/10.1007/s13394-013-0116-1>
- McNeill, K. L., & Krajcik, J. (2008). Scientific explanations: Characterizing and evaluating the effects of teachers’ instructional practices on student learning. *Journal of Research in Science Teaching*, 45(1), 533–578.
- Noviyanti, N. I., Mukti, W. R., Yuliskurniawati, I. D., Mahanal, S., & Zubaidah, S. (2019). Students’ scientific argumentation skills based on differences in academic ability. *The International Seminar on Bioscience and Biological Education*. <https://doi.org/10.1088/1742-6596/1241/1/012034>
- Nurramadhani, A. (2017). Argument-driven inquiry (ADI): The way to develop Junior High School student’s argumentation skills in science learning. *International Conference on Mathematics and Science Education*, 57(ICMSEd 2016), 128–132.
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 994–1020. <https://doi.org/10.1002/tea.20035>
- Paul, R., & Elder, L. (2014). Critical thinking: Intellectual standards essential to reasoning well within every domain of human thought. *Journal of Developmental Education*, 37(3).
- Rauf, R. A. A., Rasul, M. S., Mansor, A. N., Othman, Z., & Lyndon, N. (2013). Inculcation of science process skills in a science classroom. *Asian Social Science*, 9(8), 47–57. <https://doi.org/10.5539/ass.v9n8p47>
- Sampson, V., & Blanchard, M. R. (2012). Science teachers and scientific argumentation: Trends in views and

- practice. *Journal of Research in Science Teaching*, 49(9), 1122–1148.
<https://doi.org/10.1002/tea.21037>
- Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, 92(3), 447–472.
<https://doi.org/10.1002/sce.20276>
- Sampson, V., & Gleim, L. (2009). Argument-driven inquiry to promote the understanding of important concepts & practices in biology. *The American Biology Teacher*, 71(8), 465–472.
- Schleigh, S. (2014). Assessments in the arguments: Round robin journaling and white boarding provide multiple opportunities for assessing students and evaluating instruction. In *Science and Children*. Scilinks.
- Tsai, C., Lin, C., Shih, W., & Wu, P. (2015). The effect of online argumentation upon students' pseudoscientific beliefs. *Computers & Education*, 80, 187–197.
<https://doi.org/10.1016/j.compedu.2014.08.018>
- Walter, J. G., & Barros, T. (2011). Students build mathematical theory: semantic warrants in argumentation. *Educ Stud Math*, 78, 323–342.
<https://doi.org/10.1007/s10649-011-9326-1>
- Wattanakasiwich, P., Taleab, P., Sharma, M. D., & Johnston, I. D. (2013). Development and implementation of a conceptual survey in thermodynamics. *International Journal of Innovation in Science and Mathematics Education*, 21(1), 29–53.