check for updates

Authentic assessment development science to assess student competency

Muh. Tawil[1+]
Muhammad Amin Said[2]
Kemala Suryansari[3]

[1]State University of Makassar, Indonesia.
Email: muh.tawil@unm.ac.id
[2]FKIP Muhammadiyah University Makassar, Indonesia.
Email: muhamin@unismuh.ac.id
[3]Senior High School 1 Gowa, Indonesia.
Email: kemala1671@gmail.com

(+ Corresponding author)

ABSTRACT

This study aimed to examine authentic models of science assessment in assessing the competence of senior high school students who met the criteria of validity, practicality, and effectiveness. The methodology applied was an evaluation and developmental research. Research was conducted in senior high school in accordance with the needs of the development of authentic Statistical Analysis System (SAS) models. The participants were students of class X senior high school Gowa in South Sulawesi academic year 2020/2021. The instruments comprised (1) SAS model validation sheet; (2) science achievement test; (3) questionnaires; and (4) observation sheet. The findings of the quality authentic science assessment model met the validity and effectiveness criteria for high category learning outcomes with positive student responses. The practicality implications of the SAS model were also very worthy of applied science in teaching science for assessing the competence of senior high school students.

Contribution/Originality: The contribution of this research is having developed Statistical Analysis System (SAS) as an authentic science assessment model for qualified products that meet the criteria of validity, effectiveness and practicality.

1. INTRODUCTION

Teachers view assessment as continuous, interrelated and strongly integral part of learning process. Paper and pencil tests are the only evaluation tools carried out in learning so that all student competencies cannot be measured (AlHouli & Al-Khayatt, 2020; Kurniawati & Sukardiyono, 2018; Mohamed & Lebar, 2017; Mongkuo & Mongkuo, 2017; Susani, 2018; Suwartono & Riyani, 2019). If they are authentic assessment methods, they can measure all student competencies (Inayah, Komariah, & Nasir, 2019; Moria, Refnaldi, & Zaim, 2017; Setyawarno & Kurniawati, 2018).

2. THEORETICAL FRAMEWORK

2.1. Authentic Assessments

Authentic assessments are made in complex situations (Centoni & Antonello, 2021; Moria et al., 2017; Salirawati, 2021; Suarimbawa, Marhaeni, & Suprianti, 2017; Sutarto & Jaedun, 2018; Wiethe-Korprich & Sandra, 2017). Authentic assessment of educational achievement directly measures actual performance in the subject area. A

few studies (Fynn & Elias, 2022; Sholihah, 2021; Sutarto & Jaedun, 2018) show how students' assessment is done. Students can demonstrate that complex skills are difficult to be assessed through paper and pencil tests and require authentic assessments techniques (Aiken, 2013), as shown in Table 1. A framework of organizing authentic assessments is shown in Figure 1 based on Carin (2017), and Agus and Suprianti (2017).

Table 1. Techniques of authentic assessments.

| Complex skills that require authentic assessments techniques	
Planning and carrying out experiments	Creating graph data
Writing stories, essays, poetry	Creating the chart
Giving oral reports	Noting the observations
Being a peer tutor	Asking question
Creating a journal	Improvise
Interviewing informants	Creating computer programs
Being a local travel guide	Searching for information in the library and etc.
Correspond with science authors	Completing the literature review

Source: Aiken (2013).
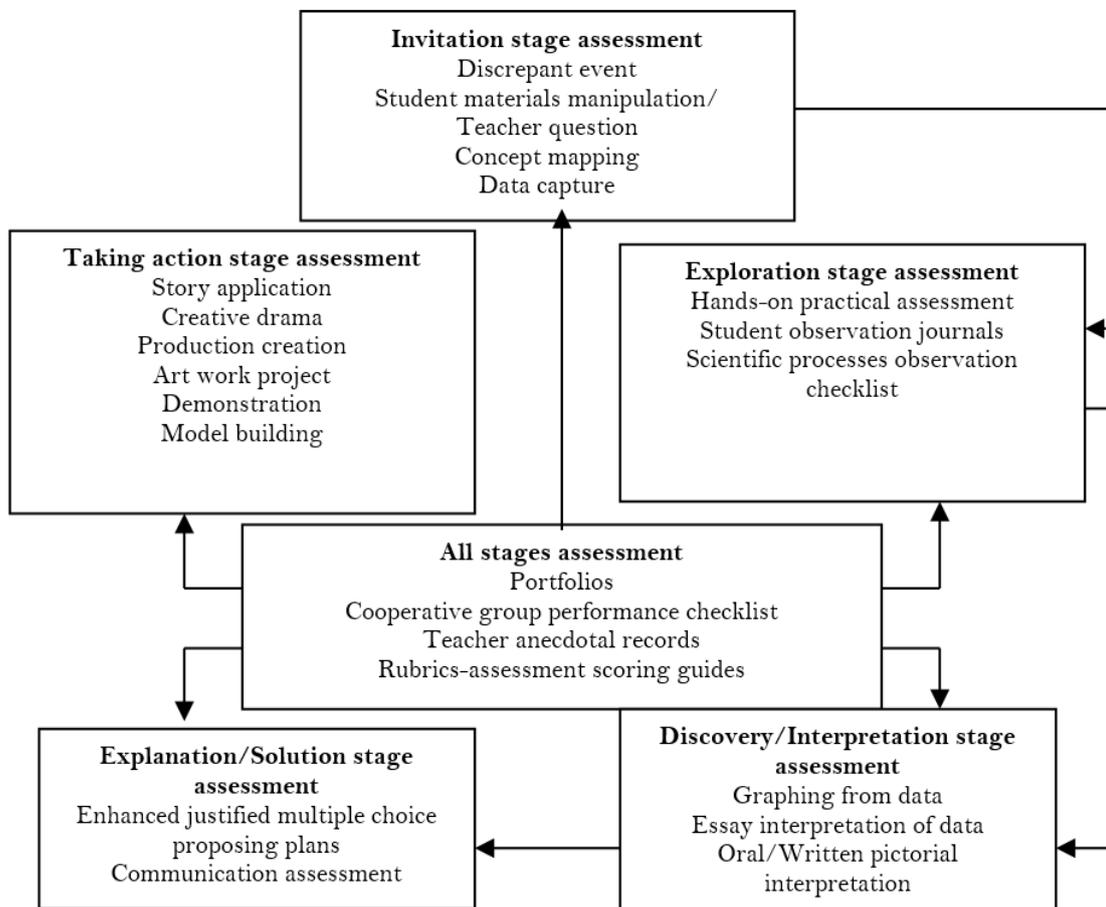
Figure 1. Authentic assessments organization in the learning cycle stage.
Source: Carin (2017) and Agus and Suprianti (2017).

2.2. Implementation of Authentic Assessment

Implementation of authentic assessment at each stage can be described in following four stages (Kurniawati & Sukardiyono, 2018):

1. *Invitational Stage*

Authentic assessments aimed to obtain information about the initial understanding of the students on a particular topic or concept. The technique used, among others, included asking questions orally or in teacher-

oriented writing to provoke students' opinions, and presenting photos or illustrations that demonstrate a process or a particular situation.

2. *Exploration Stage*

In this stage, assessment aims to explore the ability of science process skills in students. Assessment techniques may be hands-on/minds-on activity including the assessment of student activities and student observation journal.

3. *Explanation Stage*

In this stage, authentic assessment is done on the competence of students in observation, analysis, and communication.

4. *Action Stage*

This is the final stage when assessment aims to uncover the student's ability to apply his understanding of a concept in other situations. The resulting product can be student demonstrations, role playing, oral or written communications, handicrafts, and others.

3. METHOD

3.1. Type of Research

This research is categorized as developmental research, as it contributes to the development of authentic assessment models or Statistical Analysis System (SAS) models and instruments required. SAS has been accepted and recognized as a software system for data analysis and report writing. It helps to store data values and retrieve them, modify data, compute simple and complex statistical analyses, and create reports. In the development research, SAS models are qualified in determining validity, practically, and effectiveness.

3.2. Research Sites

This research was conducted at a senior high school in Gowa in South Sulawesi, in accordance with the needs of development research and application of SAS models.

3.3. Participants

The subjects were the students of class X senior high school at Gowa in South Sulawesi, in the academic year of 2020/2021. The sample size of research subjects in the test phase was 350 students who participated in a questionnaire survey.

3.4. Development Operational

Operationally the development of SAS models is done simultaneously, so when the validity of the model SAS criteria is not met, revising the devices (partially or completely) is done simultaneously with the help of SAS models support devices and related instruments. Simultaneously, the plan of teaching and performance assessment tasks and rubrics are also prepared. Operational development of authentic models of science assessment is shown in Figure 2.
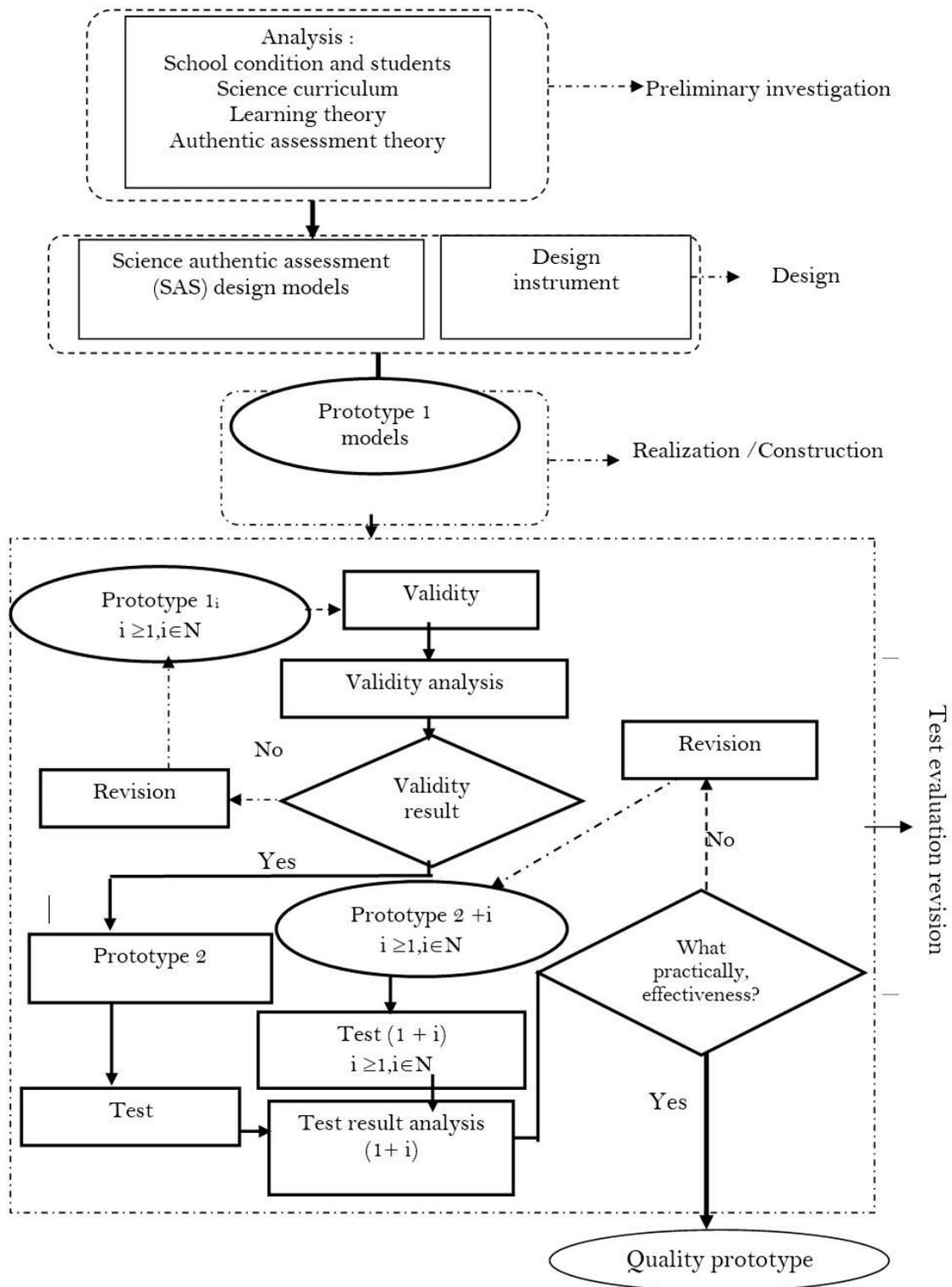
Figure 2. Development operational SAS models.

Remarks:

☐ : Activity process   ◇ : Requirement results

⬚ : Development operational   ⟶ : Main flow

⬭ : Activity results   – · ➛ : Needs if cyclic process

3.5. Instrument

The instruments consisted of (1) SAS models validation sheet; (2) science achievement test; (3) questionnaire; and 4) the observation sheet. These instruments were validated by five experts associated with the evaluation and validation of the content of the construct validation. The tests showed that these instruments were eligible for use in research.

3.6. Data Analysis

Data analysis included quality principles like validity, practically, and effectiveness as recommended in (Huizinga, Handelzalts, Nieveen, & Voogt, 2014) and Veugen, Gulikers, and Brok (2021), as shown in Table 2.

Table 2. Quality principles.

Quality aspects		
Validity	**Practicality**	**Effectiveness**
Intended (Idea+ Formal)	Consistency between	Consistency between
-State of the art knowledge	-Intended ⟷ Perceived	Intended ⟷ Experiential
-Internally consistent	- Intended ⟷ Operational	Intended ⟷ Attained

The criteria for the categorization of quality SAS models were taken from Bloom, (Fauzan, Plomp, & Gravemeijer, 2013) with the following steps.

a. Calculate average results of the validation of formula $k_i = \dfrac{\sum_{j=1}^{n} V_{ji}}{n}$ with k_i = average, the results

b. Calculate average of formula $A_i = \dfrac{\sum_{j=1}^{n} k_{ij}}{n}$

k_{ij}: Average sub-aspect to j

n: The number aspect sub-aspects of the aspects of the i-th

c. Calculate for a mean total (VR) with the formula $VR = \dfrac{\sum_{j=1}^{n} A_i}{n}$,

Table 3 presents the Validity of the SAS model (Chakraborty, Dann, Mandal, Dann, & Paul, 2021).

Table 3. Validity of the SAS model (Chakraborty et al., 2021).

Interval score	Validity criteria
$4 \leq VR \leq 5$	Very high
$3 \leq VR < 4$	High
$2 \leq VR < 3$	Less
$1 \leq VR < 2$	Low

Note: VR is the average total of the results of the assessment of experts, practitioners, and observers.

A data analysis and students' response to the implementation of supporting components of SAS models and data from questionnaires determine the positive response, meaning that students support, feel happy, and are interested in the components and the learning process through the implementation of SAS models. A negative response otherwise determines the achievement of the objectives of the activity implementation of SAS models in

terms of the students' responses. This is the degree of achievement of competence is the percentage of student mastery of the content and performance of the basic competencies of 80% of the basic competencies that have been set can be achieved. According to (Wong, Bajwa, & Fienup, 2022), a program of teaching and learning activities otherwise is very effective, to reach benchmark reference value set beforehand. Since the matter refers to the indicators, the values obtained by the students representing the student's mastery of the indicator is represented by the question, so that the control indicators can be expressed as the total value obtained by the students.

The criteria to achieve mastery learning with application of SAS models are a minimum score 80 (maximum 100) and interval score determining the level of student mastery (LSM) as categorized in Table 4.

Table 4. Criteria for level students' mastery (LSM).

Interval score	Category
90≤ LSM ≤ 100	Very high
80≤ LSM< 90	High
70≤ LSM< 80	Less
60≤ LSM<70	Low
0≤ LSM< 60	Very low

Note: LSM is the average of the results of the level of student mastery of experts, and practitioners.

4. RESULTS

4.1. Needs Analysis

As said earlier, parents and the community at large want information on the progress of the learning progress of students from time to time and comprehensively, namely the development in cognitive, affective, and psychomotor aspects. This trend indicates that the conventional, marking system "paper and pencil test" (written test), conducted in schools during this time should be time takes a form of alternative assessment and a complementary system of written test. Written tests cannot assess all students' abilities, on the other hand people have a tendency to get information about cognitive, affective, and psychomotor aspects Svidzinskaya, Baskin, and Mezentseva (2019) and Miller-Young, Marin, Poth, Vargas-Madriz, and Xiao (2021), as by applying the assessment of these aspects, they wish to measure student's attitude and competence. The assessment system that can address these needs of community is the SAS model, that has a philosophy of continuous assessment. The philosophy of this model states that if people want a complete profile of students, they must be photographed from all sides, not only in terms of cognitive, but affective, and psychomotor aspects as well. These three aspects are absolutely developed simultaneously, as required in the curriculum of 2013.

The results of this survey also found some interesting information. First, in general, people wish that the assessment criteria used in the school should be communicated to students. It means that the assessment criteria used in assessing performance results (assignments, papers, daily tests) in school should be delivered to students. Second, the public generally believe that students need to be included in assessing performance outcomes, it means that the community wants that students need to be trained to assess themselves against the results of their performance. Therefore, being an assessor of their own ability, students will be able to determine the "weakness" and "strength", so that in the end they can have the ability to overcome the problems faced in everyday life. This is in accordance with the opinion of AlHouli and Al-Khayatt (2020) and Syaifuddin (2020) which stated that by applying the self-assessment techniques, student skill competencies can be measured.

4.2. Curriculum Analysis

The results of the analysis of the structure of the content was developed based on the contents of Curriculum 2013, in particular. It was found that the 2013 science curriculum senior high school was no longer the elaboration of indicators. It meant that teachers were required to create indicators and specific learning objectives in the

curriculum. Based on this information, it is in the development of learning tools described indicators and objectives specific to each Lesson Plan, and in each of the learning plan comes with authentic assessment sheet. The task analysis was conducted to identify the stages of completion of tasks in accordance with the material. The tasks implemented on SAS models, in the form of task performance which consists of duty to report the results of observations or experiments, in general, students do well, it can be seen from the result of the ability of process skills which the average value of students.

In curriculum tests, the tasks of the competency test in general which students can do, were seen from the average results study on cognitive aspects. The tasks of students assessing themselves against the results of self-assessment of its performance and its participation in the group were also seen in the learning outcomes in the affective aspects. Similarly, data from the questionnaire results in the affective aspect of learning outcomes showed that the results were consistent with the observations and the results of the students' statements within the SAS models, all of which showed average values. It was proved that the learning outcomes of students in a very affective aspects are reliable. That is, there is a match between the observer and the statement of the results of a respondent.

In the selection task, and the task of reflection also on the subject matter students generally perform well, but the results rather than the student's performance is not analyzed further. The results of the performance of these tasks is the result of group work done at home. However, that does not mean it has no effect on the motivation of the students, because the value of the group is taken as a reference in awarding the prize to the group. This scenario is delivered to students, so that they are motivated to do the work of the group, both in completing observation or experiment, and perform tasks in the home. The results of the analysis of the information indicates that some of the concepts of temperature and heat that previously have not been mastered by students were based on the results of the initial tests (on material temperature and heat in Junior high school). Students at the time had participated in the implementation of the model SAS and these concepts were already controlled, which is reflected in the value of the average learning results obtained during the implementation of SAS models.

The results obtained from the analysis of the concept of teacher responses indicate that 100 percent of teachers said that the analysis of this concept was very helpful in implementing SAS models. It means that the analysis of the concept was practically implemented in SAS models.

4.3. SAS Models: Validity Results

Data content of this study included the validation results supporting theories SAS models development and validation constructs include constructs linkages between components of SAS models in the category of valid and fit for use. The results of the validation content and construct models of the SAS are shown in Table 5 and 6.

Table 5. Data of validity results module science authentic assessment content.

No	Rated aspect	Average aspect*
1.	Supporting theory model of science authentic assessment	4.1
2.	Step-by-step implementation model of science authentic assessment	3.9
3.	The social system in the implementation model of science authentic assessment	3.9
4.	The principle of the model management reaction science authentic assessment	4.0
5.	SAS models support system implementation	3.9
6.	Impact of the implementation and impact of the model companion science authentic assessment	4.0
7.	Implementation of the model of science authentic assessment	4.1
8.	The learning environment and management duties science authentic assessment models	3.8
9.	Science authentic assessment models evaluation	3.9
Average total = 3.9		

Note: *Average aspect is the average of the results of validity module in all science authentic assessment content.

Table 6. Data of validity results construct components science authentic assessment model.

No	Rated aspects	Average aspects*
1.	Science authentic assessment model components	3.9
2.	Science authentic assessment models support the theory	4.1
3.	Step-by-step implementation of science authentic assessment models	3.8
4.	The social system in the implementation of science authentic assessment models	3.8
5.	Reaction principle management science authentic assessment models	3.9
6.	Science authentic assessment models	3.8
7.	The implementation bridesmaids science authentic assessment models	3.6
8.	Implementation of science authentic assessment models	4.1
9.	The learning environment and management duties models science authentic assessment	3.7
10.	Science authentic assessment model evaluation	4.3
	The average total (VR) = 4.2	

Note: *Average aspects is the average of the results of validity construct components science authentic assessment model.

In general, states that the link expert contents of the components of SAS models are interrelated and do not find any contradiction between what the contents of the book with the device SAS models. Comments and suggestions from the expert including SAS models guidebook packed with modern learning theory is very relevant to the central theme of research. The handbook should be made more compact, attractive and avoid some mistakes topography. Data validation results of SAS model development support device are presented in Table 7: planning for learning; Table 8: task authentic assessment; Table 9: validity assessment authentic supplement, and Table 10: guidelines for teachers (teachers manual.

Table 7. Data of validity results for planning of learning.

No.	Rate aspects	Average aspects*
1	Indicator	3.9
2	The contents of the subject matter	4.0
3	Language	3.8
4	Time	3.9
5	Teaching methods	3.7
6	Learning closed	3.9
	Average total (VR) = 3.8	

Note: *Average aspect is the average of the results of planning for learning.

Table 8. Data of validity results for task authentic assessment.

No	Rate aspects	Average aspects*
1	Organization	4.0
2	Procedure	3.7
3	Question/problem	3.7
	Average total (VR) = 3.9	

Note: *Average aspect is average of the results of task authentic assessment

Table 9. Data of validity results for authentic assessment supplement.

No.	Rate aspects	Average aspects*
1	Organization	3.9
2	Procedure	3.9
3	Question/Problem	3.9
	Average total (VR) = 3.9	

Note: *Average aspects is the average of the results of authentic assessment supplement.

Table 10. Data of validity results for teacher manual.

No	Rate aspects	Average aspects*
1.	Introduction	3.9
2.	Science problem representation	4.2
3.	Learning method	3.8
4.	Learning closed	3.9
Average total (VR) = 3.9		

Note: *Average aspects is the average of the results of teacher manual.

Science achievement tests consisted of a test product, scientific knowledge tests, and performance tests. Each test consisted of 24 items of products initially, but after testing one item was found invalid, so there were 23 items. These 23 items represented all the material temperature and heat with the level of reliability 0.99. The performance test of 10 items, after testing produced was 0.72, within the prescribed limits.

4.4. SAS Models: Effectiveness Results

Table 11 shows the results of teacher feedback on the effectiveness of the implementation of SAS models. The assessment scores to test effectiveness were: 1 very low; 2 low; 3 sufficient; 4 high; 5 very high. Table 12 presents data effectiveness results on student book of supporting devices.

Table 11. Teacher feedback on the effectiveness of the implementation of SAS models in student manual.

No	Rate aspects	Average aspects*
1	Sub organization concept	3.6
2	Translation of troubleshooting steps temperature and heat	3.9
3	Activity	4.4
4	Learning closed	4.2
Average total (VR) = 3.9		

Note: *Average aspects is the average of the results of effectiveness of the implementation of SAS models in student manual.

Table 12. Data of effectiveness on student book of supporting devices.

No	Rate aspects	Average aspects*
1	Translation of material	3.9
2	Construction	3.9
3	Exercise	4.0
Average total (VR) = 3.9		

Note: *Average aspects is the average of the on student book of supporting devices

Table 13 presents data on the effectiveness of the authentic assessment science models.

Table 13. Data of effectiveness results of the authentic assessment science models.

No	Rate aspects	Average aspects*
1	Student learning outcomes	4.1
2	Student response against components and implementation Process science authentic assessment models	4.1
Average total (VR)= 4.1		

Note: *Average aspects is the average of the results of effectiveness science authentic assessment models.

4.5. SAS Models: Practicality Results

Table 14 shows the practicality of the model based on the observation of the implementation of SAS models support device.

Table 14. Data of practicality results of implementation of SAS models support device.

No	Rate aspects	Average aspects*
1	Measures implementation of science authentic assessment models	3.1
2	Social systems implementation of science authentic assessment models	3.9
3	Reaction principle management science authentic assessment models	3.8
IO value or value average total (VR) = 3.6		

Note: IO is intended ⟺ Operational.
*Average aspects is the average of the results of implementation of supporting device.

Science achievement test scores found that the level of student mastery (LSM) of the subject matter of science is high. Students' response to the implementation of science authentic assessment models is a slight positive, according to the teacher's response to the ease of implementing science authentic assessment models is positive.

Table 15 presents students' comment on questions related to the implementation of SAS models support device, viz.,

1. Are you able to understand the language used in the Student book, Student Worksheet, Student Manual, and SAS supplements?
2. Are you interested in the appearance (text, large font, image, layout images, colors) Student book, Student Worksheet, Student Manual, and SAS supplements?

Table 15. Data of students' comment on questions related to the implementation of SAS models.

No.	Comment	
	First question	Second question
1	Yes, because all of a language is very easy to understand	Yes, because in the student book, student worksheet, student manual, and SAS supplements, material is written clearly, accompanied by images that can allow me to learn.
2	Yes, the language of each material is clear and simple.	Yes, because the appearance makes us more interested in reading and with their images with easy to understand explanations.
3	Yes, because the language used is standard; sentences are short and clear so easy to understand.	Yes, because with the article with pictures easier for students to understand the intent rather than the text.
4	Yes, I can understand the language used in the material as it appealed to my heart	Yes, because if the appearance of text the learning atmosphere was less boring and we were happy to do our work.

Table 16 shows the results of the responses of six education experts regarding enforceability of SAS models theoretically in the field. Educational experts assessed the enforceability of SAS models in the classroom by using assessment scores: 1 very low; 2 low; 3 sufficient; 4 high; 5 very high.

Table 16. Expert response data regarding the implementation of SAS model.

No	Aspects observed and judged	Average aspects*
I	SAS models implementation procedures	4.2
2	Social aspects of SAS models implementation	4.2
3	Principal reaction management od SAS models	4.3
Average total = 4.2		

Note: *Average aspects is the average of the results of expert responses regarding the implementation of SAS model.

5. DISCUSSION AND CONCLUSION

5.1. Discussion

The analysis of the validity, effectiveness and practicality found that SAS models are supported through valid criteria applied in assessing students' competence. Student competency can be assessed in a comprehensive and sustainable manner. This finding is consistent with that of Muho and Taraj (2022) who also found formative

assessment practices helping in the sustainability of students' motivation in learning English. Student learning outcomes include cognitive aspects (products and processes), affective, and psychomotor aspects obtained from the authentic assessment. The results indicated that students meet the standards of competence specified achievement. This is because all the planning was done well. Similarly, the results obtained from the student learning achievement test at the end of the implementation of the model showed that SAS models met all the criteria. Chi, Xiu, and Zuhao (2021) conducted a study related to comparing student science and had found similar results

The results of student and teacher responses to the components and implementation of SAS models showed a positive response. It showed that in the aspect of student and teacher responses to the components and implementation of SAS models meet the criteria of effectiveness. That is, the entire device of SAS models supported the implementation of learning in the classroom and expected that students will be more motivated to learn science. Based on the results of this questionnaire, it was found that teachers followed the necessary and feasible SAS models developed on the subject of how to do more with the device making training and accompanied by the provision of laboratory equipment in accordance with the required media in the implementation of the SAS models. Similarly, the teachers stated that the strategy of implementation of SAS models can be used as a primary strategy in learning science. Vahidnia, Behzad, and Hesamoddin (2021) and Fauzan et al. (2013) also developed and validated an instrument of student attitudes towards teachers to find differences in the level of student attitudes.

The results of this study are very appropriate response to the comments of experts based on their knowledge and experience theoretically. It proved empirically SAS model implementation in the field which is derived from the observation of two (2) observers. The results indicated that the average yield assessment scores of both observers are in the high category and reliable. Thus, it can be said that the implementation of SAS models is theoretically supported by empirical data in the field with high reliability. The implementation of SAS models although observed by different people and different conditions gave inconsistent results. Toma (2021) developed a similar instrument for psychometric evaluation and got similar results. Teachers initially still face obstacles in implementing SAS models, because as a model system SAS is a form of assessment and learning system innovation. As an innovation model, the SAS model requires a shift in perspective from teachers, community, including a shift in perspective parents. Changing the way, it is not an easy task, but requires hard work and commitment. Teachers are required to be able to pay attention to individual students, monitor progress, encourage them more activity, collecting every student work to be comments, and so forth. All it requires extra time and effort. However, teachers lack strong motivation in their profession, it would be difficult to implement this SAS models. This is not in line with the results of Chi et al. (2021), Matsumoto-Royo and Ramírez-Montoya (2021) and Wong et al. (2022).

In the next phase, the teachers have been able to overcome the obstacles perceived in the early stages. Teachers, giving students the opportunity to assess themselves to the task with a friend based on the authentic assessment guidelines. Teachers keep checking on the results of such self-assessment. The impact of self-assessment is that it will foster honesty of students in doing the job. SAS Models have some weaknesses in the implementation of the class: (1) due to the limitations of the researcher, only two observers observed social skills, psychomotor skills of students, and enforceability of SAS models. This is done with the consideration that the use of many observers could affect the process of implementation of SAS models in the classroom; (2) observation of social skills, and psychomotor skills was performed in all groups. This is done with the consideration that social skills and psychomotor skills a person cannot be sampled for other people. All information to assess social skills and psychomotor skills, as well as observations were made only with one group in each meeting. There were at least seven meetings to assess the implementation of SAS models. The disadvantage of this observation is the observation frequency is too numerous but the benefit is that all students can be accessed.

The results of students in the high category and the response of students and teachers on the implementation of positive SAS models. This is in accordance with the opinion of Wiethe-Korprich and Sandra (2017) and Toma

(2021); Suwartono and Riyani (2019) which states that by applying the assessment, students' knowledge competence can be measured (Setyawarno & Kurniawati, 2018) and Sutarto and Jaedun (2018).

To conclude, this study resulted in proving SAS models as qualified products that meet the criteria of validity, effectiveness and practicality.

REFERENCES

Agus, K., & Suprianti, G. A. P. (2017). An analysis of authentic assessment implementation based on curriculum 2013 in SMP Country 4 Singaraja. *Journal of Education Research and Evaluation, 1*(1), 38-45. https://doi.org/10.23887/jere.v1i1.9551

Aiken, L. R. (2013). *Psychological tests and evaluation.* Mexico: Pearson.

AlHouli, A. I., & Al-Khayatt, A. K. A. (2020). Assessing the soft skills needs of teacher education students. *International Journal of Education and Practice, 8*(3), 416-431. https://doi.org/10.18488/journal.61.2020.83.416.431

Carin, A. A. (2017). *Teaching science through discovery* (8th ed.). New Jersey: Prentice Hall.

Centoni, M., & Antonello, M. (2021). Students' evaluation of academic courses: An exploratory analysis to an Italian case study. *Studies in Educational Evaluation, 70,* 101054. https://doi.org/10.1016/j.stueduc.2021.101054

Chakraborty, S., Dann, C., Mandal, A., Dann, B., & Paul, M. (2021). Effects of rubric quality on marker variation in higher education. *Studies in Educational Evaluation, 70,* 100997. https://doi.org/10.1016/j.stueduc.2021.100997

Chi, S., Xiu, F. L., & Zuhao, W. (2021). Comparing student science performance between hands-on and traditional item types: A many-facet Rasch analysis. *Studies in Educational Evaluation, 70,* 100998. https://doi.org/10.1016/j.stueduc.2021.100998

Fauzan, A., Plomp, T., & Gravemeijer, K. (2013). The development of an RME-based geometrycourse for Indonesian Primary schools. In T. Plomp, & N. Nieveen (Eds.), Educational design research - Part B. Illustrative cases. In (pp. 159-178). SLO: Netherlands Institute for Curriculum Development.

Fynn, A., & Elias, O. M. (2022). Continuous online assessment at a South African open distance and e-learning institution. *Frontiers in Education Original Research, 27*(79), 1-13. https://doi.org/10.3389/feduc.2022.791271

Huizinga, T., Handelzalts, A., Nieveen, N., & Voogt, J. M. (2014). Teacher involvement in curriculum design: Need for support to enhance teachers' design expertise. *Journal of Curriculum Studies, 46*(1), 33-57. https://doi.org/10.1080/00220272.2013.834077

Inayah, N., Komariah, E., & Nasir, A. (2019). The practice of authentic assessment in an EFL speaking classroom. *Studies in English Language and Education, 6*(1), 152-162. https://doi.org/10.24815/siele.v6i1.13069

Kurniawati, A., & Sukardiyono, S. (2018). The development of authentic assessment instrument to measure science process skill and achievement based on students' performance. *Journal of Physics Education Research & Development, 4*(2), 65-74. https://doi.org/10.21009/1.04203

Matsumoto-Royo, K., & Ramírez-Montoya, S. M. (2021). Core practices in practice-based teacher education: A systematic literature review of its teaching and assessment process. *Studies in Educational Evaluation, 70,* 101047. https://doi.org/10.1016/j.stueduc.2021.101047

Miller-Young, J., Marin, L. F., Poth, C., Vargas-Madriz, L. F., & Xiao, J. (2021). The development and psychometric properties of an educational development impact questionnaire. *Studies in Educational Evaluation, 70,* 101058. https://doi.org/10.1016/j.stueduc.2021.101058

Mohamed, R., & Lebar, O. (2017). Authentic assessment in assessing higher order thinking skills. *International Journal of Academic Research in Business and Social Sciences, 7*(2), 466-476.

Mongkuo, M. Y., & Mongkuo, M. Y. (2017). Testing the factorial equivalence of the collegiate learning assessment performation task diagnostic instrument across lower class and upper class predominantly black college students. *International Journal of Education and Practice, 5*(6), 95-103. https://doi.org/10.18488/Journal.61.2017.56.95.103

Moria, E., Refnaldi, R., & Zaim, M. (2017). *Using authentic assessment to better facilitate teaching and learning: The case for students' writing assessment.* Paper presented at the Sixth International Conference on Languages and Arts (ICLA 2017). Atlantis Press.

Muho, A., & Taraj, G. (2022). Impact of formative assessment practices on student motivation for learning the English language. *International Journal of Education and Practice, 10*(1), 25-41. https://doi.org/10.18488/61.v10i1.2842

Salirawati, D. (2021). Authentic assessment in the pandemic period. *Journal of The Indonesian Society of Integrated Chemistry, 13*(1), 21-31.

Setyawarno, D., & Kurniawati, A. (2018). Implementation of authentic assessment in science learning at Indonesian schools. *Journal of Science Education Research, 2*(2), 47-55.

Sholihah, M. (2021). Authentic assessment in online learning during the covid-19 pandemic. *Edunesia: Scientific Journal of Education, 2*(2), 576-580.

Suarimbawa, K. A., Marhaeni, A., & Suprianti, G. (2017). An analysis of authentic assessment implementation based on curriculum 2013 in SMP Country 4 Singaraja. *Journal of Education Research and Evaluation, 1*(1), 38-45. https://doi.org/10.23887/jere.v1i1.9551

Susani, R. (2018). The implementation of authentic assessment in extensive reading. *International Journal of Education, 11*(1), 87-92.

Sutarto, H. P., & Jaedun, M. P. D. (2018). Authentic assessment competence of building construction teachers in indonesian vocational schools. *Journal of Technical Education and Training, 10*(1).

Suwartono, T., & Riyani, C. (2019). Authentic assessment in ELT: Hopes, challenges, and practices. *Reflection on Education: Scientific Journal of Education, 9*(2).

Svidzinskaya, G. B., Baskin, Y. G., & Mezentseva, M. E. e. (2019). Using the Semantic Differential Method to Assess the Learning Motivation and Attitude of First-Year Students towards Chemistry in University of EMERCOM of Russia. *International Journal of Education and Practice, 7*(2), 88-100.

Syaifuddin, M. (2020). Implementation of authentic assessment on mathematics teaching: Study on junior high school teachers. *European Journal of Educational Research, 9*(4), 1491-1502. https://doi.org/10.12973/eu-jer.9.4.1491

Toma, B. R. (2021). Measuring children's perceived cost of school science: Instrument development and psychometric evaluation. *Studies in Educational Evaluation, 70*, 101009. https://doi.org/10.1016/j.stueduc.2021.101009

Vahidnia, F., Behzad, G. H., & Hesamoddin, S. (2021). Development and validation of students' attitudes towards teacher's pet phenomenon scale in the higher education setting: Differences by levels of study and grade-point-average. *Studies in Educational Evaluation, 70*, 101000. https://doi.org/10.1016/j.stueduc.2021.101000

Veugen, M. J., Gulikers, J. T. M., P, & Brok, D. (2021). We agree on what we see: Teacher and student perceptions of formative assessment practice. *Studies in Educational Evaluation, 70*, 101027. https://doi.org/10.1016/j.stueduc.2021.101027

Wiethe-Korprich, M., & Sandra, B. (2017). Prospective educators as consumers of empirical research: An authentic assessmentapproach to make their competencies visible. *Empirical Research in Vocational Education and Training, 9*(8), 2-26. https://doi.org/10.1186/s40461

Wong, K. K., Bajwa, T., & Fienup, D. M. (2022). The application of mastery criterion to individual operants and the effects on acquisition and maintenance of responses. *Journal of Behavioral Education, 31*(3), 461–483. https://doi.org/10.1007/s10864-020-09420-3