DEVELOPMENT OF THREE TYPES OF ASSESSMENT IN LEARNING MATHEMATICS ON THE SYSTEM OF LINEAR EQUATIONS WITH THREE VARIABLES

Ishomah Nadiyah[1], Ali Shodikin[2*]

[1,2]Mathematics Education Study Program, Faculty of Natural Sciences and Mathematics,
Surabaya State University
e-mail: alishodikin@unesa.ac.id

Abstract

This study aims to develop three types of assessments in mathematics learning for the Three Variable Linear Equation System material: initial, formative, and summative assessments. These three assessments are designed for use on a digital *platform* to enhance the efficiency of implementation and student involvement in the learning process. The study uses a *Research and Development* (*R&D*) approach with the *ADDIE model* consisting of five stages: analysis, design, development, implementation, and evaluation. The initial assessment in the form of a learning style questionnaire is distributed through *Typeform* to identify student learning characteristics. The formative assessment is designed in the form of an interactive quiz using the *Nearpod application* which includes multiple-choice questions. In the meantime, the summative evaluation is created as a context-based exam that gauges the *knowledge*, *application*, and *reasoning* dimensions. Four experts used *Aiken's V test* to validate the developed instrument, and *Alpha Cronbach's* reliability test was used to assess its dependability. With an *Aiken's V* > 0.80, the validation results demonstrate the instrument's high level of validity. The reliability test yields a moderately consistent *Alpha Cronbach* value of 0.45. The implementation's outcomes demonstrate that this technology-based evaluation can help identify student traits early on, offer feedback throughout the learning process, and measure learning success in a comprehensive way. Thus, creating this assessment could be a useful substitute for enhancing students' understanding of the Three Variable Linear Equation System.

Keywords: Assessment, Three Variable Linear Equation System, Validity, Reliability, Technology-Based Assessment

BACKGROUND

In order to support the efficacy of mathematics learning, assessment is crucial. These days, assessment serves more purposes than just gauging outcomes; it also serves to track the learning process and modify focused learning tactics (Cusi et al., 2023) . Teachers can better understand students' learning needs and developmental stages by implementing well-designed assessments. As a result, assessment plays a crucial role in the development and application of adaptive learning.

One of the more challenging subjects in mathematics is the Three Variable Linear Equation System, which frequently presents challenges for students. This is due to the Three Variable Linear Equation System's simultaneous requirement for modelling abilities, algebraic manipulation, and logical reasoning (Dewi et al., 2020) . Many students struggle to choose the best solution approach and to translate contextual problem representations into a system of

---

equations (Kusuma et al., 2024) . This issue shows that an assessment tool that can pinpoint particular learning challenges is required.

It has been demonstrated that alternative teaching strategies, like the *crossing method*, help students better comprehend the Three-Variable Linear Equation System (Deogratias, 2014) . Comprehensive assessment support for this type of approach is still lacking, though. Assessments that are only oriented toward the final result are not enough to capture students' thinking processes and problem-solving strategies. Therefore, an assessment is needed that can reveal students' abilities more comprehensively through various stages of learning.

Advances in educational technology open up new opportunities for implementing more interactive and efficient assessments. Digital *platforms such as Typeform* and *Nearpod* allow assessments to be conducted online, provide immediate feedback, and store learning outcome data more systematically (Weigand et al., 2024) . Cevikbas, Kaiser, and Schukajlow (2024) stated that digital assessments can increase student participation and motivation because they are flexible and adaptive to individual needs.

However, the development of a good assessment cannot be separated from testing its validity and reliability. Validity ensures that the assessment measures the intended competency, while reliability shows the consistency of assessment results in different applications (Aydoğan & Tanrıöğen, 2024) . Additionally, a number of studies stress that unvalidated tools run the risk of giving inaccurate information when it comes to learning decision-making (Brochey-Taylor & Taylor, 2024) .

A thorough picture of students' learning progress can be obtained by combining the three forms of assessment: initial, formative, and summative. According to Levy-Feldman (2025) summative assessments show students' final performance, formative assessments track the learning process, and initial assessments can assist teachers in understanding students' readiness for learning. Teachers can implement more focused learning activities with this tiered assessment approach.

One of the main topics in the study of current trends in mathematics education is the incorporation of technology with assessment techniques (Cevikbas et al., 2024) . In addition to supporting a more individualised learning process, technology-based assessments facilitate *real-time* analysis of learning outcomes by teachers. Additionally, this integration creates room for creativity in creating tests that suit the needs and characteristics of students.

Based on the description above, this study focuses on the development of three types of technology-based assessments for learning Three-Variable Linear Equation Systems. The assessment instrument was developed using the *ADDIE model* and tested through content validity and internal reliability. Thus, this study is expected to contribute to providing valid, reliable, and contextual assessments to improve the quality of mathematics learning at the secondary level.

METHOD

Types and Design of Research

This type of research is development research or *Research and Development* (*R&D*) using the *ADDIE model* , which consists of five stages: *Analysis, Design, Development, Implementation, and Evaluation.* (Alniza Azman et al., 2024) & (Ergin & Diker Coskun, 2024) . This model was chosen to produce valid, reliable, and applicable assessment products in mathematics learning, especially in the Three Variable Linear Equation System material.

Time and Place of Research

This research was conducted in the 2024/2025 academic year at one of the high schools in Surabaya City, namely SMA Dharma Wanita Surabaya located at Jl. Kendangsari Surabaya. The location selection was carried out purposively by considering the school's readiness to implement digital-based learning and the availability of adequate internet access.

Research Subject

The subjects in this study were 20 grade X students who were studying the material of Three Variable Linear Equation Systems. The subject selection technique used the *purposive sampling method* to ensure the representation of student characteristics and technological readiness.

Research Procedures

The research procedure follows the five stages *of ADDIE* . The analysis stage is carried out by studying the curriculum, learning conditions, and assessment needs of the Three Variable Linear Equation System material. Initial, formative, and summative assessment formats are prepared during the design phase. Instrument creation and expert validation using validation sheets are part of the development stage. The implementation stage is carried out by testing the instrument on grade X students. Finally, the evaluation stage includes analysis of validation data, reliability, and effectiveness of assessments in the learning process.

Research Instruments

The research instruments used consisted of three types of assessments. First, the initial assessment was in the form of a learning style questionnaire distributed through *the Typeform platform* . Second, the formative assessment was in the form of an interactive *Nearpod-based quiz* containing multiple choices. Third, the summative assessment was in the form of a context-based test that measured three aspects: *knowing* , *applying* , and *reasoning* ((Cusi et al., 2023) **;** (Weigand et al., 2024)).

Data Analysis Techniques

Data analysis techniques consist of two parts: validity analysis and reliability analysis. Content validity is analyzed using *Aiken's V formula* based on the assessment of four validators. Internal reliability was analyzed using *Cronbach's Alpha* to assess consistency between items. Interrater reliability tests such as in the study by Takahashi et al. (2021) are an important reference in ensuring consistency of results, even though the contexts are different. In addition, student assessment data were analyzed descriptively to see the achievement of learning objectives and the effectiveness of the instruments developed.

Success Indicators

Indicators of success in this study include (1) *Aiken's V value* > 0.80 as an indicator of high validity, (2) *Cronbach's Alpha value* of at least 0.40 as acceptable moderate reliability for exploratory instruments, and (3) active student involvement in digital assessment as an indicator of the effectiveness of instrument implementation ((Mohd Razali et al., 2025) **;** (Levy-Feldman, 2025)).

RESULTS AND DISCUSSION

Research result

This study aims to develop an assessment of mathematics learning on the material of Three Variable Linear Equation Systems. The results were obtained from three types of assessments used, namely non-cognitive initial assessment, formative assessment, and summative assessment. Each type of assessment provides different information regarding student characteristics, engagement, and learning outcomes.

Non-Cognitive Initial Assessment Results

Initial assessments were conducted to identify students' learning styles . The results of filling out the questionnaire distributed through *Typeform* showed that most students had visual (60%) and kinesthetic (20%) learning styles , while the rest were included in the auditory category (20%). This data is important as a basis for teachers to design learning strategies that are in accordance with the characteristics of the majority of students.

Table 1 Distribution of Student Learning Styles Based on Initial Assessment

Learning Styles	Number of Students	Percentage
Visual	12	60%
Auditory	4	20%
Kinesthetic	4	20%

Table 1 shows that the majority of students (60%) tend to absorb information through visual displays. This is the basis for developing learning strategies and presenting assessments that involve diagrams, colors, and other visualizations.

Formative Assessment Results

Formative assessments are conducted through the *Nearpod platform* and involve interactive activities such as matching answers, card games, and contextual questions consisting of 5 questions. The results of the formative assessment show that most students already understand the basic concept of the Three Variable Linear Equation System.
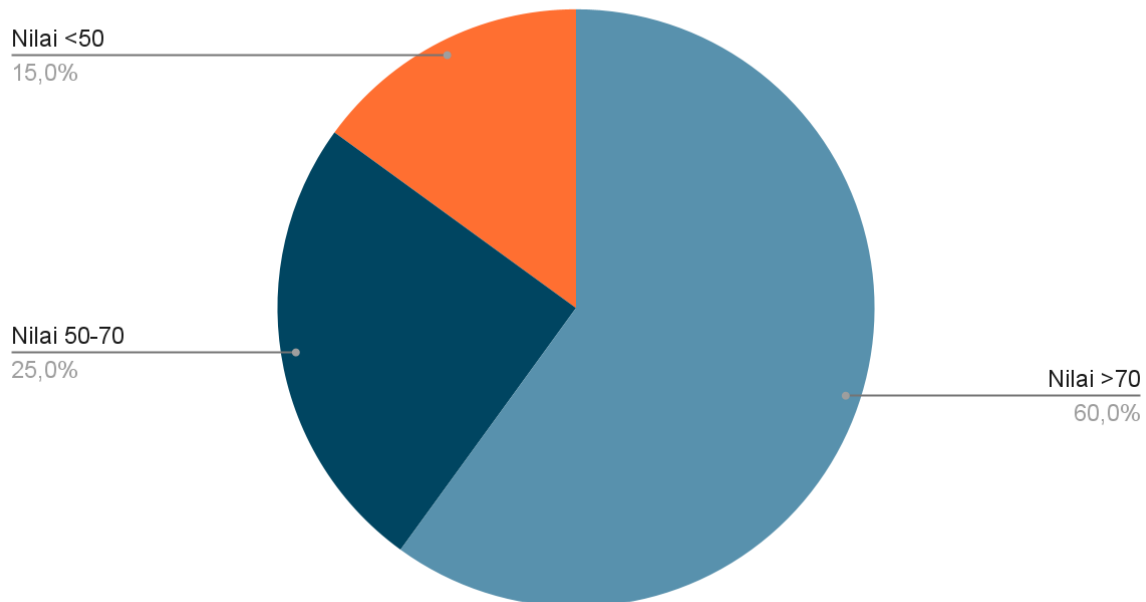
Figure 1 Percentage of Student Scores on Nearpod Formative Assessment

Figure 1 shows that 60% of students scored above 70, indicating very good mastery of the material. Meanwhile, 25% were in the range of 50–70 and another 15% were below 50, indicating that learning reinforcement was still needed. These results indicate that most students have understood the material well during the learning process.

Summative Assessment Results
a. Trial Results
    1. Validity of Summative Assessment

Validitas berdasarkan Validator pada Kriteria A											
Butir	**Validator**				**S1**	**S2**	**S3**	**S4**	**∑s**	**n(c-1)**	**V**	**Ket**
	I	II	III	IV								
Butir 1	4	4	3	3	3	3	2	2	10	12	0,83333	Sangat Tinggi
Butir 2	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 3	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 4	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 5	4	1	4	2	3	0	3	1	7	12	0,58333	Sedang
Butir	**Validator**				**S1**	**S2**	**S3**	**S4**	**∑s**	**n(c-1)**	**V**	**Ket**
	I	II	III	IV								
Butir 1-5	20	17	19	17	15	12	14	12	53	60	0,88333	Sangat Tinggi

Figure 2 Validity Test Results on Criterion A

Based on the assessment for variable A, namely "questions by the objectives and indicators of the assessment" by the four validators, the results obtained were that out of 5 questions, there were 4 questions that were classified as "very high" and 1 question in the "moderate" category, so the researcher retained 5 questions to take data from respondents by reprocessing the words according to the indicators and objectives of the assessment according to the

suggestions of the validators. Thus, in variable A , which contains 5 questions, it is said to be valid to take data from 20 respondents.

Butir	Validator				S1	S2	S3	S4	∑s	n(c-1)	V	Ket
	I	II	III	IV								
Butir 1	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 2	4	4	3	3	3	3	2	2	10	12	0,83333	Sangat Tinggi
Butir 3	4	3	4	3	3	2	3	2	10	12	0,83333	Sangat Tinggi
Butir 4	4	3	4	3	3	2	3	2	10	12	0,83333	Sangat Tinggi
Butir 5	1	1	1	1	0	0	0	0	0	12	0	Sangat Rendah
Butir	Validator				S1	S2	S3	S4	∑s	n(c-1)	V	Ket
	I	II	III	IV								
Butir 1-5	17	15	16	14	12	10	11	9	42	60	0,7	Tinggi

Validitas berdasarkan Validator pada Kriteria B

Figure 3 Results of Validity Test on Criterion B

Based on the assessment for variable B, namely "questions presented by the dimensions of knowledge and dimensions of cognitive processes" by the four validators, the results obtained were that out of 5 questions, there were 4 questions that were classified as "very high" and 1 question in the "very low" category, so that the researcher improved 1 question to take data from respondents by presenting it by the dimensions of knowledge and cognitive processes according to suggestions from the validators. Thus, in variable B there is 1 question that still needs improvement so that it can be said to be valid for taking data from 20 respondents.

Butir	Validator				S1	S2	S3	S4	∑s	n(c-1)	V	Ket
	I	II	III	IV								
Butir 1	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 2	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 3	4	3	4	3	3	2	3	2	10	12	0,83333	Sangat Tinggi
Butir 4	3	3	4	4	2	2	3	3	10	12	0,83333	Sangat Tinggi
Butir 5	2	1	2	1	1	0	1	0	2	12	0,16667	Sangat Rendah
Butir	Validator				S1	S2	S3	S4	∑s	n(c-1)	V	Ket
	I	II	III	IV								
Butir 1-5	17	15	18	16	12	10	13	11	46	60	0,76667	Tinggi

Validitas berdasarkan Validator pada Kriteria C

Figure 4 Results of Validity Test on Criterion C

Based on the assessment for variable C, namely "graphs, tables, pictures, maps, symbols or the like are presented clearly, functionally" by the four validators, the results obtained were that from 5 questions, there were 4 questions that were classified as "very high" and 1 question in the "very low" category, so that the researcher improved 1 question to collect data from respondents by presenting pictures or the like clearly according to the suggestions of the validators. Thus, in variable C there is 1 question that still needs improvement so that it can be said to be valid for collecting data from 20 respondents.

Validitas berdasarkan Validator pada Kriteria D											
Butir	Validator				S1	S2	S3	S4	∑s	n(c-1)	V	Ket
	I	II	III	IV								
Butir 1	3	3	3	3	2	2	2	2	8	12	0,66667	Tinggi
Butir 2	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 3	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 4	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 5	2	3	2	3	1	2	1	2	6	12	0,5	Sedang
Butir	Validator				S1	S2	S3	S4	∑s	n(c-1)	V	Ket
	I	II	III	IV								
Butir 1-5	17	18	17	18	12	13	12	13	50	60	0,83333	Sangat Tinggi

Figure 5 Validity Test Results on Criterion D

Based on the assessment for variable D, namely "the formulation of commands/questions requires logical and argumentative answers (in essay questions)" by the four validators, the results obtained were that from 5 questions, there was 1 question that was classified as "high", 3 questions were classified as "very high", and 1 question was in the "moderate" category, so the researcher retained 5 questions to collect data from respondents by reprocessing the words that required respondents to answer logically and submit their arguments according to the suggestions of the validators. Thus, variable D, which contains 5 questions, is said to be valid to collect data from 20 respondents.

Validitas berdasarkan Validator pada Kriteria E											
Butir	Validator				S1	S2	S3	S4	∑s	n(c-1)	V	Ket
	I	II	III	IV								
Butir 1	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 2	4	3	3	4	3	2	2	3	10	12	0,83333	Sangat Tinggi
Butir 3	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 4	3	4	4	3	2	3	3	2	10	12	0,83333	Sangat Tinggi
Butir 5	2	2	3	3	1	1	2	2	6	12	0,5	Sedang
Butir	Validator				S1	S2	S3	S4	∑s	n(c-1)	V	Ket
	I	II	III	IV								
Butir 1-5	17	17	18	18	12	12	13	13	50	60	0,83333	Sangat Tinggi

Figure 6 Results of Validity Test on Criterion E

Based on the assessment for variable E, namely "logical answer key reasons and by the truth of the concept" by the four validators, the results obtained were that from 5 questions, there were 4 questions that were classified as "very high" and 1 question in the "moderate" category, so that the researcher retained 5 questions to take data from respondents by reprocessing the words according to the concept as suggested by the validators. Thus, in variable E, which contains 5 questions, it is said to be valid to take data from 20 respondents.

Validitas berdasarkan Validator pada Kriteria F											

Butir	Validator				S1	S2	S3	S4	∑s	n(c-1)	V	Ket
	I	II	III	IV								
Butir 1	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 2	4	4	3	3	3	3	2	2	10	12	0,83333	Sangat Tinggi
Butir 3	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 4	4	4	4	4	3	3	3	3	12	12	1	Sangat Tinggi
Butir 5	3	3	3	3	2	2	2	2	8	12	0,66667	Tinggi

Butir	Validator				S1	S2	S3	S4	∑s	n(c-1)	V	Ket
	I	II	III	IV								
Butir 1-5	19	19	18	18	14	14	13	13	54	60	0,9	Sangat Tinggi

Figure 7 Results of Validity Test on Criterion F

Based on the assessment for variable F, namely "the language used is clear and appropriate for high school students so that it is easy for students to understand" by the four validators, the results obtained were that from 5 questions, there were 4 questions that were classified as "very high" and 1 question in the "high" category, so the researcher retained 5 questions to take data from respondents by reprocessing the words using clear language and easy for high school students to understand according to suggestions from the validators. Thus, in the F variable, which contains 5 questions, it is said to be valid to take data from 20 respondents.

2. Reliability of Summative Assessment

	butir soal	butir soal	butir soal	butir soal	butir soal	skor total
siswa 1	2	2	1	1	2	8
siswa 2	2	3	1	0	0	6
siswa 3	2	1	1	1	1	6
siswa 4	2	3	2	1	2	10
siswa 5	2	2	1	1	3	9
siswa 6	1	2	1	1	0	5
siswa 7	1	1	1	1	1	5
siswa 8	1	3	3	1	3	11
siswa 9	1	2	1	2	0	6
siswa 10	1	2	1	1	0	5
siswa 11	1	3	3	1	2	10
siswa 12	1	3	3	2	2	11
siswa 13	1	3	3	1	2	10
siswa 14	2	1	2	1	1	7
siswa 15	2	1	2	1	1	7
siswa 16	1	2	1	0	0	4
siswa 17	1	2	1	1	1	6
siswa 18	1	3	1	0	3	8
siswa 19	1	1	1	1	2	6
siswa 20	2	1	1	1	2	7
Varians Total						4,6605
Varians Butir Soal	0,2526	0,6816	0,6816	0,2605	1,0947	2,9711
Nilai Cronbach Alpha						0,4531

$$\alpha = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2}\right)$$

Dengan α = koefisien reliabilitas instrumen; k = banyaknya butir pertanyaan dalam instrumen; $\sum \sigma_i^2$ = jumlah varians butir instrumen; σ_t^2 = varians skor total.

Figure 8 Results of Reliability Test Using *EXCEL*

In the image above, the results of the reliability analysis test of questions using *EXCEL* are seen from the *Alpha value* . The results of the reliability test on the questions presented in the table with an alpha value equal to 0.45313 are included in the "sufficient" criteria. This can be interpreted that the questions have sufficient consistency in measuring students' abilities.

b. Item Analysis
   1. Level of Difficulty of Question Items

PEDAMATH
Journal on Pedagogical Mathematics

Volume 7, No. 2, Tahun 2025, pp. 50-65
Ishomah Nadiyah, dkk.

Figure 9 Results of Analysis of the Level of Difficulty of Question Items

The image above shows the results of the analysis of the level of difficulty of the question items. Based on calculations, the level of difficulty of the five questions is categorized as "moderate". This indicates that the questions have a level of difficulty that is appropriate to the average student's ability, not too easy and not too difficult.

2. Distinction Power of Question Items

Figure 10 Results of the Analysis of the Differential Power of Question Items

The image above shows the results of the item discrimination power analysis. Based on the analysis, the items that have sufficient, good, and very good discrimination power are found in items 2, 3, and 5, which means the items are difficult for the lower group but very easy for the upper group. Then the poor discrimination power is found in items 1 and 4, which means the items are difficult for the upper group.

Discussion

The results of the initial assessment showed that most students have a visual learning style (60%). This finding is in accordance with the characteristics of high school students who

PEDAMATH
Journal on Pedagogical Mathematics

Volume 7, No. 2, Tahun 2025, pp. 50-65
Ishomah Nadiyah, dkk.

generally find it easier to understand material through visual displays such as pictures, colors, and diagrams. This condition is in line with the view of Weigand et al. (2024) which states that visual-based assessment and learning can increase information retention and the appeal of learning materials. Therefore, the use of graphic-based media in assessments is very appropriate for students with a tendency towards a visual learning style .

In addition, the implementation of formative assessment through *the Nearpod platform* provides an illustration that 60% of students achieved scores above 70. This shows that most students have understood the concept of the Three-Variable Linear Equation System well during the learning process. Saarela's study (2024) highlights the importance of understanding the solution structure of three-variable equations as a foundation for compiling mathematical models correctly. The implementation of interactive formative assessments can encourage active student involvement, as stated by Cusi et al. (2023) that technology-based formative assessments can increase students' attention and response to the material being studied. However, the group of students who scored below 50 (15%) showed the need for further differentiation and mentoring efforts for certain groups.

The summative assessment developed consists of contextual questions that measure the dimensions of *knowing* , *applying* , and *reasoning* . The results of student achievement indicate that summative assessment is able to comprehensively evaluate students' understanding of the Three-Variable Linear Equation System in a real context. This finding is in line with Levy-Feldman's (2025) opinion which states that context-based summative assessment is able to describe student achievement holistically, not just procedural memorization.

According to Santos-Trigo (2024) , problem-based assessment not only assesses the final results but also trains students to think reflectively, which is very important in understanding the Three-Variable Linear Equation System. This can be seen from the fairly high *reasoning achievement in the summative assessment.* On the other hand , the impact of learning on students' interests and self-confidence also needs to be considered. These results are in line with Zhu & Kaiser (2022) who stated that appropriate teaching practices can improve students' self-efficacy and mathematics test achievement.

From the aspect of instrument quality, *Aiken's V value* for the assessment shows higher validity (>0.80) . The validity aspect in instrument development is also highlighted in the research of Schweidtmann et al. (2022) which emphasizes the importance of data suitability to the analysis model. This is in line with the findings of Cetin (2024) which emphasizes that reliable assessment instruments must go through a validation and reliability process to ensure consistency of assessment, especially in the context of learning observation. This indicates that the contents of the instrument are in accordance with the competency indicators and learning context. High validity also strengthens that the instrument has passed the expert testing stage carefully, as stated by Aydoğan and Tanrıöğen (2024) that content validity is very important in developing assessments that are in accordance with student needs and curriculum objectives.

The summative assessment, on the other hand, received a value of 0.45 from the reliability test using *Alpha Cronbach*, falling into the sufficient category. This value is still appropriate for an exploratory instrument that is being used in its early stages of development according to Mohd Razali et al. (2025) . This reliability value indicates that this instrument can be used with a note of further improvement even though the consistency between the items is not very high.

A high *Aiken's V* value suggests that every assessment was created with consideration for how well the content matched the learning indicators. The result of Schweidtmann et al. (2022) who highlighted the significance of validity limits in data-based instrument models, further support this validity. Furthermore, Takahashi et al. (2021) study on inter-rater reliability in the medical field supports the notion that reliability falling into the sufficient category is still appropriate in the context of early development.

Lastly, the analysis of this data demonstrates that a thorough picture of students' learning processes and results can be obtained by integrating three different assessment types: formative, summative, and initial. As stressed by Cetin (2024) in creating a scale of teacher observation skills, this finding supports the significance of teachers' observation abilities in thoroughly evaluating students' learning processes. The findings of Zhu & Kaiser (2022) show that appropriate teaching practices have a direct impact on students' learning interests, self-efficacy, and mathematics test results. Assessment strategies like this support students' needs-based learning while providing a valid evaluation of their learning outcomes **(** (Bakker et al., 2021) **;** (Chorlay et al., 2022) **)** . In other words , the development of validated digital-based assessments can be an alternative solution to strengthen mathematics learning, especially in the Three-Variable Linear Equation System material.

CONCLUSION

This study produces three types of assessments in learning the Three Variable Linear Equation System, namely initial, formative, and summative assessments based on digital. The assessments have been validated by experts and show a high level of validity and sufficient reliability to be used in the context of class X of high school.

Initial assessment successfully identified students' learning styles as a basis for designing appropriate learning. Formative assessment has been proven to help students during the learning process through the provision of interactive feedback. Meanwhile, summative assessment can measure students' final achievements in the aspects of *knowing* , *applying* , and *reasoning* through contextual questions.

In general, the integration of these three types of assessments contributes to a more adaptive and responsive learning process to students' needs . The use of digital *platforms such as Typeform* and *Nearpod* also strengthens the efficiency of implementation and student involvement in the evaluation process.

Based on the results of this study, it is recommended that mathematics teachers begin to integrate the three types of assessments in their learning practices, especially for complex materials such as the Three Variable Equation System. In addition, further research can be conducted with a wider scope or the development of more diverse instruments to improve the quality and consistency of technology-based assessments in the future.

THANK-YOU NOTE

BIBLIOGRAPHY

Alniza Azman, Nor, Mohd Isa Hamzah, Khadijah Abdul Razak, and Hafizhah Zulkifli. 2024. *Digital Competence Among Islamic Teachers: A Pilot Study on Validity and Reliability*. Vol. 23.

Aydoğan, Süleyman Sırrı, and Zeynep Meral Tanrıöğen. 2024. "School Corporate Social Responsibility Scale: A Validity and Reliability Study." *Shanlax International Journal of Education* 12(4):34–48. doi: 10.34293/education.v12i4.7539.

Bakker, Arthur, Jinfa Cai, and Linda Zenger. 2021. "Future Themes of Mathematics Education Research: An International Survey before and during the Pandemic." *Educational Studies in Mathematics* 107(1). doi: 10.1007/s10649-021-10049-w.

Brochey-Taylor, Julia, and Joseph A. Taylor. 2024. "Educational Research and Reviews Synthesizing Validity and Reliability Evidence for the Draw a Scientist Test." 19(3):44–52. doi: 10.5897/ERR2023.4385.

Cetin, Abdullah. 2024. "Teacher Observation Skills Scale: Validity and Reliability." *International Journal of Contemporary Educational Research* 11(3):363–75. doi: 10.52380/ijcer.2024.11.3.692.

Cevikbas, Mustafa, Gabriele Kaiser, and Stanislaw Schukajlow. 2024. "Trends in Mathematics Education and Insights from a Meta-Review and Bibliometric Analysis of Review Studies." *ZDM - Mathematics Education* 56(2):165–88. doi: 10.1007/s11858-024-01587-7.

Chorlay, Renaud, Kathleen Michelle Clark, and Constantinos Tzanakis. 2022. "History of Mathematics in Mathematics Education: Recent Developments in the Field." *ZDM - Mathematics Education* 54(7):1407–20. doi: 10.1007/s11858-022-01442-7.

Cusi, Annalisa, Florian Schacht, Gilles Aldon, and Osama Swidan. 2023. "Assessment in Mathematics: A Study on Teachers' Practices in Times of Pandemic." *ZDM - Mathematics Education* 55(1):221–33. doi: 10.1007/s11858-022-01395-x.

Deogratias, Emmanuel. 2014. "ISSN: 2148-225X." *International Online Journal of Education and Teaching (IOJET)* 2022(2):1022–31.

Dewi, Irmawati Liliana Kusuma, Zaenuri, Dwijanto, and Mulyono. 2020. "Identification of Mathematics Prospective Teachers' Conceptual Understanding in Determining Solutions of Linear Equation Systems." *European Journal of Educational Research* 10(3):1157–70. doi: 10.12973/EU-JER.10.3.1157.

Ergin, Atilla, and Yelkin Diker Coskun. 2024. "Design Thinking Scale Development: Assessing Reliability and Validity." *International Journal on Social and Education Sciences* 6(3):319–33. doi: 10.46328/ijonses.670.

Kusuma, A. P., St Budi Waluya, and S. Mariani. 2024. "Algebraic Thinking Process of Students with High Mathematical Ability in Solving Linear Equations Based on Cognitive Systems." *Pegem Journal of Education and Instruction* 14(3):146–59. doi: 10.47750/pegegog.14.03.14.

Levy-Feldman, Irit. 2025. "The Role of Assessment in Improving Education and Promoting

Educational Equity." *Education Sciences* 15(2). doi: 10.3390/educsci15020224.

Mohd Razali, Mohd Norlizam, Aida Hanim A. Hamid, Bity Salwana Alias, and Azlin Norhaini Mansor. 2025. "Validity and Reliability: Instruments of Teacher Competency in the Context of Small Schools in Peninsular Malaysia." *Journal of Education and Learning (EduLearn)* 19(1):487–94. doi: 10.11591/edulearn.v19i1.21574.

Saarela, Aleksi. 2024. "On the Solution Sets of Three-Variable Word Equations." *Theory of Computing Systems*. doi: 10.1007/s00224-024-10193-9.

Santos-Trigo, Manuel. 2024. "Problem Solving in Mathematics Education: Tracing Its Foundations and Current Research-Practice Trends." *ZDM - Mathematics Education* 56(2):211–22. doi: 10.1007/s11858-024-01578-8.

Schweidtmann, Artur M., Jana M. Weber, Christian Wende, Linus Netze, and Alexander Mitsos. 2022. "Obey Validity Limits of Data-Driven Models through Topological Data Analysis and One-Class Classification." *Optimization and Engineering* 23(2):855–76. doi: 10.1007/s11081-021-09608-0.

Takahashi, Yosuke, Yuji Fujino, Kohei Miura, Ayumi Toida, Tadamitsu Matsuda, and Shigeru Makita. 2021. "Intra- and Inter-Rater Reliability of Rectus Femoris Muscle Thickness Measured Using Ultrasonography in Healthy Individuals." *Ultrasound Journal* 13(1). doi: 10.1186/s13089-021-00224-8.

Weigand, Hans Georg, Jana Trgalova, and Michal Tabach. 2024. "Mathematics Teaching, Learning, and Assessment in the Digital Age." *ZDM - Mathematics Education* 56(4):525–41. doi: 10.1007/s11858-024-01612-9.

Zhu, Yan, and Gabriele Kaiser. 2022. "Impacts of Classroom Teaching Practices on Students' Mathematics Learning Interest, Mathematics Self-Efficacy and Mathematics Test Achievements: A Secondary Analysis of Shanghai Data from the International Video Study Global Teaching InSights." *ZDM - Mathematics Education* 54(3):581–93. doi: 10.1007/s11858-022-01343-9.