

Development of Linear Motion Assessment Using Evidence-Centered Design

Hepi Kharisda Hulu¹, Judyanto Sirait^{1*}, Firdaus¹, Nuri Balta²

¹Department of Physics Education, Faculty of Teacher Training and Education, Tanjungpura University, Pontianak, Indonesia

²Suleyman Demirel University, Kaskelen, Kazakhstan

*Email: judyanto.sirait@fkip.untan.ac.id

Article History:

Received date: December 9 2024

Received in revised from: February 27 2025

Accepted date: March 4 2025

Available online: March 15 2025

Citation:

Hulu, H.K., Sirait, J., Firdaus & Balta, N. 2025. Development of linear motion assessment using evidence-centered design. *JUPI (Jurnal IPA dan Pembelajaran IPA)*, 9(1):49-64.

Abstract. Linear motion is one of the basic concepts in physics, so an assessment that can measure students' abilities in this material is necessary. The assessments created by teachers tend to only evaluate students' abilities in formula calculations without measuring their understanding in depth. This research aims to develop assessments on the topic of linear motion using the evidence-centered design (ECD) framework. ECD is a framework that focuses on evidence of students' abilities. ECD consists of three components: claim, evidence, and task. The process of developing this assessment uses the research and development method by following the ADDIE stages: analysis, design, development, implementation, and evaluation. Data were collected through interviews with physics teachers, analyzing teachers' documents, validating the assessment, and conducting task trials. The trials were conducted in two stages: one-to-one trials with one student and small-scale trials with 10 students. The analysis techniques included calculating validity using the Aiken Index, descriptive analysis of student responses in the one-to-one trial, and small-scale trials. The competencies measured in the assessment are based on scientific practice, including using and developing representations, analyzing and interpreting data, explaining with evidence, and evaluating information. Each measured competency in the assessment consists of two essay tasks. The validation results show that the developed assessment has high validity with an average Aiken score of 0.97. Teachers can use this assessment as a reference through stages to develop assessments that evaluate the achievement of learning objectives.

Keywords: R&D, Linear Motion, Assessment, Evidence-Centered Design, Scientific Practice

© 2025 The Authors. This open access article is distributed under a (CC-BY-SA License)



Introduction

Assessment plays a crucial role, especially in physics education. Assessments serve as a tool to gauge students' comprehension of physics concepts. Additionally, assessments serve to provide teachers and students with relevant feedback on the collaborative learning achievements (Kemendikbudristek, 2022). Assessments in physics learning should not only evaluate students' calculation skills, but also their critical thinking and problem-solving abilities (Bashoor & Supahar, 2018). Therefore, excellent assessment becomes a necessity in improving the quality of physics education and equipping students with the skills needed in the 21st century.

One of the basic and important concepts in physics is the linear motion. Kinematics of motion, one of the topics in physics education, studies the movement of an object

without considering the forces causing that motion (Giancoli, 2001). Students' understanding of kinematics concepts affects almost all concepts in physics and often becomes a source of errors in the application of more complex physics concepts (Jufriadi et al., 2023; Poutot & Blandin, 2015). Mastery of kinematic concepts will directly impact mastery of other physics concepts. A simple example is the relationship between speed and force in Newtonian mechanics, where misunderstandings in the understanding of speed can lead to misconceptions about force (Stoen et al., 2020).

Although assessments play a significant role in the learning process, the assessments given by teachers often do not fully align with the previously established learning objectives, especially in the subject of kinematics of motion. Assessments often prioritise calculations and formula usage over the comprehension of concepts. Many students succeed in calculation problems but still fail to understand the physical meaning of several kinematic motion concepts, leading to misconceptions (Rahmadani et al., 2023).

An interview with a senior school physics teacher revealed that current assessments primarily focus on memorizing formulas and simple calculations, failing to assess students' deep understanding or their ability to apply kinematics concepts in more complex situations. The analysis of teaching documents further reinforces this finding, revealing that assessments in several teaching modules used by teachers solely concentrate on formulas without taking into account the achievement of established learning goals. For instance, the linear motion questions only require the calculation of speed and acceleration. This is not enough to assess whether students truly understand the concept of kinematics or can apply it in everyday life. This issue necessitates an assessment that measures the achievement of learning objectives and provides clear evidence.

The importance of the concept of kinematics has prompted numerous studies to investigate and explore learners' understanding and difficulties with the concept (Jufriadi et al., 2023). However, research on the comprehensive and in-depth design of assessments remains limited (Bhakti et al., 2023; Pellegrino et al., 2014). For instance, Weisdiyanti & Juliani's (2023) research concentrates on the general validity, reliability, and differentiation ability of assessments without delving deeply into their design. Ellianawati et al., (2020) research reveals that many kinematics assessments still struggle to delve deeply into students' conceptual understanding. Furthermore, Jufriadi et al. (2023) revealed that despite the increased use of technology for assessments, a more systematic design is still necessary to evaluate understanding in kinematics material. Klein et al. (2017) also uncovered gaps in kinematics assessments, revealing that most assessments solely assess memory and calculation skills, neglecting to assess deeper abilities that enable the application of previously acquired knowledge in real-world scenarios. Further research is necessary to develop assessments that can comprehensively and deeply measure students' abilities.

One of the solutions related to assessment issues is using evidence-centered design (ECD). According to Mislevy et al. (2003), ECD is a systematic approach to designing assessments that focus on concrete evidence of learners' abilities. In ECD, assessments are designed to collect relevant evidence about students' abilities, so that they can show students' level of understanding (Bechard et al., 2019; Müftüolu & Sahin, 2024). Additionally, Haris et al. (2019) revealed that ECD not only assesses understanding in depth but also provides important information to teachers about learners' progress in achieving learning objectives. According to Sirait (2024), one of the advantages of ECD is that evidence of the achievement of learning objectives and learning outcomes has been identified since the planning stage, making it easier for teachers to measure students' abilities. In line with this, Bechard et al. (2019) state that ECD helps educators design more structured and systematic assessments, allowing for a more accurate evaluation of students' abilities. In designing assessments using ECD, it is important to involve the three main components: claim, evidence, and task (Mislevy & Haertel, 2007; Newton et al.,

2021; Steinkuehler, 2011). The established learning objectives establish a connection with the claims, which represent statements about the competencies under measurement. The data collected to support the claim is known as evidence, and it encompasses the knowledge, skills, and abilities necessary to achieve the learning objectives. The questions are designed to motivate learners to make connections between the claim and the available evidence, and to delve deeper into their understanding. Therefore, it is crucial to incorporate these three components into the design of assessments using ECD to effectively evaluate students' learning achievements (Mislevy et al., 2012).

In formulating learning objectives that are components of claims in ECD, it is important to integrate them with elements of scientific practice (SP). SP is an approach that emphasizes the direct involvement of learners in the scientific process (Lager & Lavonen, 2023; Laverty et al., 2016). In the context of assessment, integrating SP allows learners to engage directly in the scientific process, which not only tests their knowledge but also their ability to think critically and solve problems (Duschl & Bybee, 2014). Integrating SP in assessment provides learners with the opportunity to demonstrate their understanding of physics concepts through practical application, in line with the principles of assessment using ECD (Pellegrino et al., 2016). Thus, this research aims to develop a linear motion assessment using ECD to measure the achievement of learning objectives.

Methods

This research employs the research and development method in conjunction with the ADDIE development model, which encompasses the phases of analysis, design, development, implementation, and evaluation (Branch, 2009). The ultimate goal is to produce a linear motion assessment using ECD. This method involves a qualitative approach to analyze needs and a quantitative approach to measure the validity of the developed assessment. This research has only progressed to the development stage develop, resulting in a valid assessment that can be used to evaluate the achievement of learning objectives in motion kinematics. Figure 1 shows the steps in the development of a linear motion assessment using ECD.

This development commences with an analysis phase that encompasses gap analysis, material analysis, curriculum analysis, and needs analysis with the aim of designing assessments that are both relevant and effective. The gap analysis shows that physics assessments generally focus on the use of formulas and calculations without measuring students' conceptual understanding in depth. The student analysis adjusts the assessment to the characteristics of 11th-grade high school students, while the content analysis establishes linear motion as the main concept in this assessment. The curriculum analysis ensures that the assessment aligns with the merdeka curriculum, supporting the achievement of deep competencies.

The design stage aims to formulate the specifications of the upcoming assessment. The steps include defining assessment goals, determining the shape and number of assessments, designing the components involved in assessments development, and writing an assessment. Establishing assessment objectives is necessary in this research to ensure alignment with learning needs and the achievement of learning goals. Learning objectives are formulated based on SP which refers to science competencies (Saribas & Ceyhan, 2015; Windschilt, 2018). Next, determining the form and number of assessments. The total of questions as 8 items in essay form. Establishing the components of the assessment as a guideline is crucial when designing assessments (Henderson, 2024). The three ECD components of this assessment are claims, evidence, and task. Claims consist of: (1) learning outcomes that align with the achievements in the Merdeka Curriculum, specifically for high school physics subjects regarding linear motion, (2) learning objectives derived from the learning outcomes and using the SP approach. SP focuses on science

competencies that emphasize the scientific process experienced by students (Lager & Lavonen, 2023). The evidence component includes: (1) knowledge, skills, and abilities that students must possess to achieve learning objectives, (2) evidence, which is an important element in this assessment and shows that students have achieved the learning objectives. Task consist of three components: (1) characteristics of the task, dictate the structure, focus, and use of the questions in a specific context, (2) questions are designed to measure learning objectives, and (3) scoring guidelines are used to assess and grade students' work or performance using established criteria.

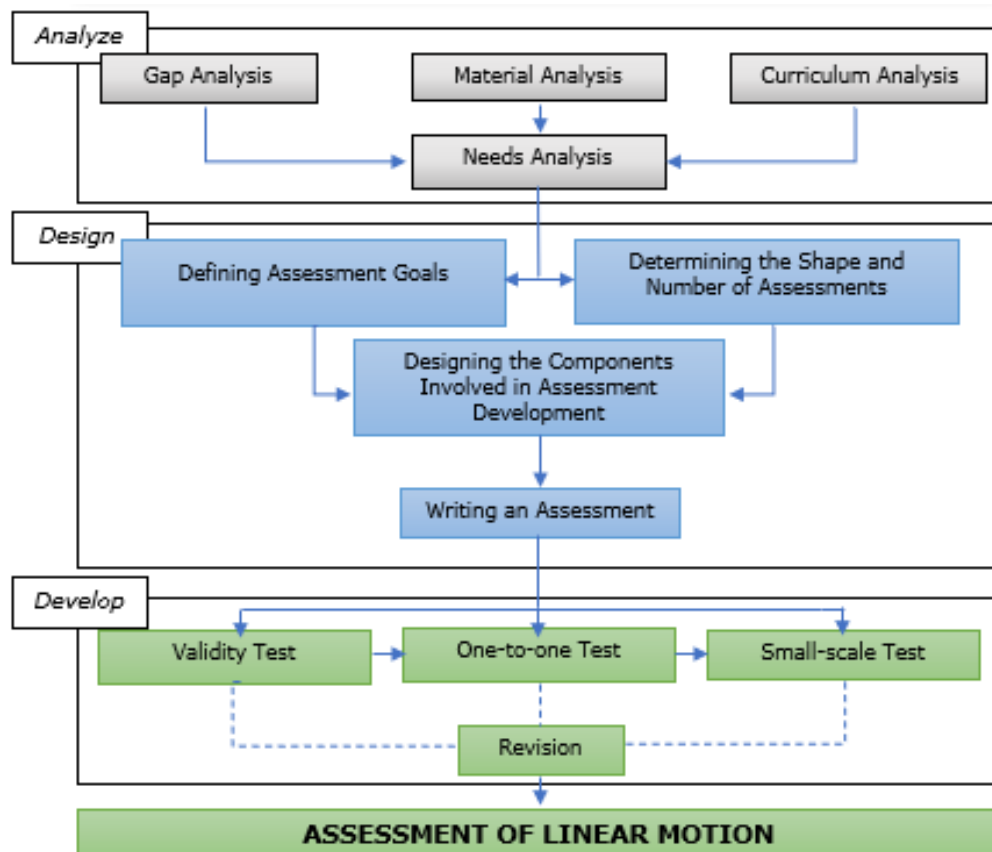


Figure 1. The flow of development of linear motion assessment using ECD

The final step in the design phase is to create the assessment itself. The assessment is made based on the components that have been created previously. By following these steps, a structured assessment can be designed to effectively evaluate students' understanding of linear motion.

In the development stage, it involves assessment validation, one-to-one testing, small-scale trials, and revisions. Three physics education lecturers and three physics teachers validated the assessment using a validation sheet, which also included suggestions and input for improvement. The determination of validity value is carried out using the Aiken index (Aiken, 1985), which can be calculated using the following formula:

$$V = \frac{\sum s}{n(c-1)}; s = r - I_0 \quad (1)$$

Following the validation stage, the assessment was revised based on the validators' suggestions and feedback. Next, one-to-one testing of the assessment is conducted to evaluate participants' responses to the developed version. Based on the results of the one-

to-one trial, further revisions are made. The next stage is a small-scale trial involving 10 students. The results of this small-scale trial were then used to make the final revisions to the assessment.

Results and Discussion

Analysis Stage

In this stage, include gap analysis, material analysis, curriculum analysis, and needs analysis. The gap analysis aims to identify constraints in the learning assessment. Data was obtained through interviews with three high school physics teachers. Based on the interviews, all three expressed that the assessments they conducted tended to focus on memorization and the use of formulas, making them less capable of measuring the achievement of learning objectives. The teacher said, "I usually create assessments without considering the learning objectives I have set beforehand. This could potentially be an oversight in my assessment creation process. Furthermore, I possess limited knowledge on how to develop effective assessments that accurately gauge the attainment of learning objectives." The module analysis further supports this finding, revealing that teachers have not yet fully designed their assessments to measure learning objectives. The ACER, UNESCO Institute for Statistics (2017) study underscores the significance of crafting unambiguous learning assessments that are in line with educational objectives, emphasizing the enhancement of learning and teaching outcomes instead of solely assessing memorization or limited understanding.

The analysis of materials plays a vital role in the development of linear motion. Kinematics assessments involve a deep understanding of basic concepts related to the movement of objects, such as position, velocity, acceleration, and the relationships between linear motion variables. The purpose of these assessments is to gauge students' proficiency in applying kinematic principles in practical situations, such as resolving issues involving distance, time, and speed calculations. The material also encompasses an understanding of both uniform linear motion and uniformly accelerated linear motion (GLBB). Using the SP, this assessment aims to measure not only students' knowledge but also their ability to apply the knowledge they already possess (Crawford & Capps, 2018). Furthermore, curriculum analysis is conducted to align the assessment with the merdeka curriculum, which establishes learning outcomes such as comprehending one-dimensional and two-dimensional motion concepts and vector analysis. Thus, this assessment is relevant to the established learning outcomes.

Finally, a needs analysis is conducted to ensure the assessment supports concept-based learning, not just rote memorization and formula application. The analysis was conducted to identify the required product type, taking into account the findings from gap, material, and curriculum analyses. Based on the analysis results, teachers require an assessment to guide them in measuring the achievement of learning objectives.

Design Stage

The design phase is the second stage in this research, which aims to develop assessment specifications. The steps taken at this stage include (1) defining assessments goals; (2) determining the shape and number of assessments; (3) designing the components involved in assessment development; and (4) writing an assessment. This stage is designed to ensure that the assessment not only adheres to quality standards but also aligns with learning needs. The following section will explain how these steps, with a systematic approach, support teachers in creating instruments that effectively evaluate various aspects of learning.

First, defining assessment goals. Teachers can use this assessment as a guide to create effective instruments that evaluate students' understanding of linear motion and

optimally measure the achievement of learning objectives. The finding that high-quality questions, like HOTS questions, can enhance students' critical and creative thinking abilities, including in understanding kinematics concepts, highlights the importance of effective assessment (Yuliantaningrum & Sunarti, 2020). Good questions must also be valid, relevant to the material, and allow for an in-depth evaluation of students' problem-solving skills (Arifin 2016; Yuliantaningrum & Sunarti, 2020).

Second, determining the shape and number of assessments. The shape of the questions used is essay questions because this type allows students to develop a deeper understanding, present arguments in a structured manner, and demonstrate critical and analytical thinking skills (Brookhart, 2010). In addition, this format allows teachers to evaluate the thinking process of students, not just the final result, thereby providing a more comprehensive insight into their conceptual understanding (Santrock, 2011). The number of questions is designed to be 8 questions that can be completed in 60 minutes. This number takes into account the cognitive load of the students and the complexity of the questions, ensuring that students have enough time to read, understand, and answer each question well (Arikunto, 2012).

Third, designing the components involved in assessment development. The assessment of linear motion is designed based on the components of evidence-centered design (ECD), which include claim, evidence, and task. The claims include components such as learning outcomes and learning objectives. Linear motion learning outcomes are the same as phase F learning outcomes for grade 11. They require students to understand one-dimensional and two-dimensional motion concepts and use vectors in their analysis. SP structure the learning objectives, encompassing seven main components to develop scientific competence. The focus on using and developing representations; analyzing and interpreting data; explaining with evidence, and evaluating information. The evidence component includes the knowledge, skills, and abilities that students must possess to achieve the learning objectives, with an important element being evidence that demonstrates success in achieving those objectives. The task component was designed to measure learning achievements. This component consists of question characteristics, questions, and scoring guidelines to ensure accurate evaluation.

The final step in this design phase is writing an assessment that aligns with the intended learning objectives. Each component of the assessment is carefully designed to measure students' understanding of the taught concepts, ensuring that their responses accurately reflect their level of competency. This step is crucial in determining whether students have successfully grasped the material and can apply their knowledge in meaningful ways. A well-structured assessment not only evaluates students' ability to recall information but also gauges their capacity to using and developing models, analyzing and interpreting data, explaining with evidence, and evaluating information critically. Additionally, the assessment can help educators identify students' learning difficulties, allowing for instructional strategies to be adjusted to enhance their understanding. Thus, assessment functions not only as an evaluation tool but also as an integral part of the continuous learning process. Table 1 provides an example of an assessment framework specifically designed to evaluate students' ability to assess and interpret information effectively, demonstrating their level of competency in this area.

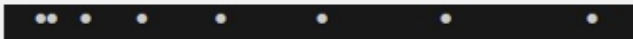
Table 1. Example of linear motion assessment components using ECD

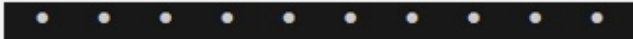
Scientific Practice	Evaluating Information
Learning Outcomes	Students are able to understand the concept of one-dimensional and two-dimensional motion along with the use of vectors and their analysis.
Learning Objective	Using ticker timer results based on experimental data, students can evaluate information about the acceleration of toy cars in a linear motion experiment.

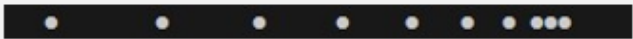
Knowledge, Skills, and Abilities	• Understanding the concept of acceleration. • The ability to analyze data from ticker timers is available. • Ability to evaluate information.
Evidence	• Students are able to correctly distinguish between constant, variable, or zero acceleration. • Students are able to write down the provided information by distinguishing between true and false information.
Characteristics of the Task	Given information about the concept of acceleration, students are able to evaluate it.

In evaluating students understanding of the materials, assessment components were designed to measure question characteristics that are crucial. The formulation of each question must reflect the intended competencies for measurement, for example, in information evaluation. Therefore, to effectively achieve the measurement objectives, it is crucial to link each type of question with the appropriate assessment characteristics. A good question is not only a test of knowledge but also the students' ability to think critically and solve problems in depth. Figure 2 is an example of a question to assess students' ability to evaluate information.

Rudy and Rika have just conducted 3 experiments related to linear motion using toy cars and a ticker timer. Here are the three recorded results of the object's trajectory on the ticker timer tape.

a) 

b) 

c) 

Based on the three experimental results presented above, Rudy and Rika discussed the acceleration of the toy car.

Rudy: "From our experiment results, I think the acceleration in picture (a) is positive because the points are getting farther apart, meaning the object's speed is increasing. In picture (b), since the distance between the points is always the same, it means the acceleration is constant. In contrast, picture (c) displays a negative acceleration, as the object slows down as the distance increases."

Rika: I agree with your explanation for images (a) and (c), but I disagree with your explanation for image (b). In my view, the acceleration in picture B is not constant but rather zero."

Rudy: "Why do you think that?"

Rika: "The object's speed remains constant if the distance along the path remains the same." If the object's speed remains constant, its acceleration is actually zero, not constant."

Rudy: Oh, do you mean that when speed is constant, acceleration is zero?"

Rika: "Exactly. Acceleration only occurs if there is a change in speed, whether it increases or decreases."

In your opinion, who is right in this discussion, Rudy or Rika? Explain your reasoning using the concept of acceleration!

Figure 2. Example of question for evaluating information

Designing a scoring guide for each student answer is necessary after creating the assessment. When creating scoring guidelines, it's crucial to clearly list each component of evidence gathered during the assessment process. This evidence includes all results obtained by students through the tasks or exams they undertake, which demonstrate their understanding and ability to apply the concepts learned. The scoring guidelines should demonstrate the objective measurement of each aspect of learning, including material understanding, analytical ability, and concept application. By linking evidence components with scoring criteria, assessments can provide relevant and accurate feedback, as well as

a clear picture of students' achievement of learning objectives. Table 2 is an example of a scoring guideline based on evidence.

Table 2. Rubric for scoring students answer in evaluating information competency.

No	Evidence	Assessment Indicator	Maximum Score
1	Students are able to correctly distinguish between constant, changing, or zero acceleration.	• Students are able to clearly differentiate between constant, changing, and zero acceleration based on ticker timer data.	3
		• Students are able to distinguish most of it correctly, but there are minor mistakes.	2
		• The students are only able to correctly distinguish one part.	1
		• Students cannot distinguish correctly.	0
2	Students are able to write down the provided information by distinguishing between true and false information.	• Students are able to write down the provided information by distinguishing between true and false information.	3
		• Students can analyze information partially correctly, but there are errors in the explanation or the explanation is incomplete.	2
		• Students can analyze inaccurate or incomplete information	1
		• There is no analysis or the analysis is completely wrong.	0
3	Students are able to write the conclusion correctly.	• Students are able to draw conclusions clearly and accurately.	3
		• The conclusion provided is partially correct, but there are errors in reasoning or it is insufficiently supported by adequate information.	2
		• The conclusion provided is inaccurate or not relevant to the available information.	1
		• No conclusion is provided or the conclusion is completely wrong.	0

The assessment of learning objectives achievement reflects the principle that evaluation should be holistic and encompass various dimensions of understanding, including the depth of concepts and the analytical abilities of the learners (Brookhart, 2013). With a clear scale, educators can identify areas that need improvement by students, in accordance with the Kemendikbud (2017) guidelines, which emphasize the importance of assessment in supporting meaningful learning. Furthermore, this approach emphasizes the significance of logical reasoning, which can enhance the attainment of learning objectives as delineated (Arends, 2012).

Development Stage

This stage represents the third phase of the assessment validation process. The initial step in this stage is to validate the assessment both quantitatively and qualitatively. According to Sudijono (2016), content validity serves as an objective parameter, ensuring the accuracy of the assessment's created content, which aims to measure learning objectives. Three panel validators from high school physics teachers conduct the first stage of content validity, while three expert validators from physics education lecturers conduct the second stage. This assessment to be validated so that it can be used as an effective measuring tool (Sirait et al., 2023a).

Based on the validation results, the validator provided several suggestions regarding the developed assessment. The panel validator suggested improvements in the use and selection of diction, both in the narrative and the question instructions, to make them clearer and easier to understand. The word "situation" is recommended to be replaced with "condition" to make the language more familiar. Additionally, the use of double questions, such as "when and where," should be avoided. It is advisable to separate those questions into two different instructions so that learners can focus more on answering without confusion. Expert validators also provided several inputs to refine the assessment. First, the context of the questions needs to be designed in such a way that the position of the objects is clearly explained, so that students can accurately understand the relationship between position, distance, and time. Second, the statements on knowledge, skills, and abilities must be aligned with the correct concepts, especially in distinguishing between distance and position, to prevent misunderstandings. Third, vector symbols need to be written consistently according to scientific notation rules, such as adding arrows to vector symbols. Lastly, it is recommended to include visual illustrations such as sketches, graphs, or tables to support the presentation of the questions and to make it easier for students to analyze the available information.

In the validation process, interviews were conducted with several physics teachers who served as panel validators in the development of this assessment. In their responses to this assessment instrument, they highlighted the novelty of reviewing the expected evidence from the students. This makes the assessment process more focused on what students are able to present, in accordance with the abilities, knowledge, and skills derived from the learning objectives. The teacher also expressed appreciation for the use of ECD, stating, "ECD is an approach that I believe is highly recommended for developing kinematics assessment instruments because it allows for assessment designs that focus on gathering evidence of students' abilities." This approach can assess students' abilities in depth by integrating various types of questions, such as problem-solving and reflection, which encompass aspects of conceptual understanding and its application in real-life situations. Additionally, the teacher assesses that the components of SP, such as developing representations, analyzing data, and evaluating information, are highly relevant to kinematics material and appropriate for the abilities of 11th-grade students. Overall, the teacher believes this assessment has great potential to uncover in-depth evidence about students' abilities in achieving the established learning objectives.

In addition to receiving improvement suggestions, the validators also assessed the content of the assessment. The scores obtained from the validators were then analyzed using the Aiken index on the aspects of content, construction, and language. The calculation results for the content aspect yielded an average Aiken index of 0.98, categorized as very good. This score indicates that the tested material aligns with the learning objectives and the chosen concepts. Next, the calculation results for the construction aspect yielded an average score of 0.98, categorized as very good. This result demonstrates the concise and clear structure of the created assessment. Then, the calculation results for the language aspect yielded an average Aiken index of 0.95, categorized as very good. This score demonstrates the assessment's use of proper and correct grammar. Based on these three aspects, the overall average Aiken index is 0.97, categorized as very good. Table 3 shows the content validity scores for each question.

Table 3. The Result of Expert's Validation

Scientific Practice	Question Number	Score	Category	Remarks
Using and Developing Models	1	0.97	Very High	Valid
	2	0.96	Very High	Valid
Analyzing and Interpreting Data	3	0.97	Very High	Valid
	4	0.98	Very High	Valid
Explaining with Evidence	5	0.98	Very High	Valid

Evaluating Information	6	0.98	Very High	Valid
	7	0.97	Very High	Valid
	8	0.98	Very High	Valid

The content validity results show that all items have a very high validity value, with an average of 0.97 and a score range between 0.96 and 0.98. All assessment items are declared valid and fall into the "very high" category (Aiken, 1985). High content validity indicates that the assessment items are significantly relevant to the learning objectives being measured, and this aligns with the principle of validity in assessment design, which emphasizes the match between the assessment and the expected competencies (Sirait et al., 2023; Widodo, et al., 2021). With very high validity, this assessment can be relied upon to measure students' ability to understand kinematics concepts accurately.

After going through the validation and revision process, the revised assessment was then tested in a one-to-one trial with one student within 60 minutes. The purpose of this trial is to evaluate students' responses to the questions that have been prepared, as well as to assess the adequacy of the time allocated for completing all the questions (Sugiyono, 2017). According to Apriyani et al. (2020), one-to-one testing is very important for identifying potential issues that may arise in the assessment, including question comprehension and the allocated working time. The trial results show that students can understand the questions well, but some questions are still considered complicated, especially those that require deeper analysis or lengthy calculations. This is due to the still insufficient understanding of linear motion. In terms of time, students feel that 60 minutes is sufficient to complete the assessment.

Based on the results of the one-to-one trial and student response questionnaires, this assessment is deemed suitable for a small-scale trial involving 10 eleventh-grade students. The purpose of this small-scale trial is to re-evaluate the effectiveness of the questions and the time required to complete them under more representative conditions (Sirait & Oktavianity, 2021). After the trial, students were given a questionnaire focusing on clarity of instructions, the use of terms in the questions, the clarity of the questions, and aspects of time adequacy. Clarity of instructions refers to the extent to which the directions in the questions can be well understood by students, so they do not encounter difficulties in following the solution steps. The use of terms in the questions assesses whether the terms used are in accordance with the concepts that have been learned and do not cause confusion. The clarity of the questions relates to how the questions are designed to be unambiguous and easily understood. The aspect of time adequacy refers to the extent to which the given time is sufficient to complete all the questions without rushing or running out of time.

The survey results indicate that students stated that the instructions in the questions were clear, the terms used were easy to understand, and the questions were presented well without ambiguities that could hinder comprehension. However, in terms of time adequacy, the survey results provided after the small-scale trial showed that 9 out of 10 students felt that the 60 minutes given was sufficient to complete all the questions. They feel that they can work on the questions carefully within the available time. However, one student expressed that the given time was not sufficient because some questions took longer to understand and complete well. The level of difficulty of the questions and the speed of each student in completing them can affect the differences in students' perceptions of the sufficiency of time. In addition to receiving feedback from students regarding the assessment that has been prepared, teachers also gain an overview of their answers. Through those answers, the teacher can analyze the extent to which the established learning objectives have been achieved. As an example, Table 4 shows the average scores obtained from a small-scale trial.

Table 4. Analysis of the scores obtained from the results of a small-scale trial

Scientific Practice	Question Number	Average Score	Maximum Score	Score obtained (%)
Using and Developing Models	1	9.9	12	82.5
	2	8.6	12	71.6
Analyzing and Interpreting Data	3	6.6	9	62.2
	4	7.9	9	87.7
Explaining with Evidence	5	10.6	12	81.5
	6	8.7	12	72.5
Evaluating Information	7	8.3	9	92.2
	8	8.2	9	91.1

Based on the Table 4, the evaluating information skill achieved the highest result, with question number 7 reaching 92.2%. This indicates that students have a very strong ability to assess and interpret information critically and draw accurate conclusions based on the given data. On the other hand, the analyzing and interpreting data skill in question number 3 obtained the lowest percentage, 62.2%, suggesting that students still struggle with effectively analyzing and interpreting data. This difference highlights the need for improvement in data analysis skills, while the ability to evaluate information is already quite strong.

This finding is also reflected in the overall achievement of learning objectives, which is categorized into three levels based on the total score obtained. Students with high scores (80–100%) are considered to have successfully achieved the learning objectives, demonstrating a strong understanding, accurate answers across all indicators, and well-structured reasoning. Those with moderate scores (50–79%) have sufficiently met the learning objectives, showing a basic understanding, though some weaknesses remain in analysis or reasoning. Meanwhile, students with low scores (0–49%) have not yet achieved the learning objectives, as their answers still contain significant conceptual errors, inaccurate analyses, and reasoning that does not fully support their understanding.

Besides score analysis from the small-scale trial, data on students' ability to answer questions was also collected, as shown in Table 5.

Table 5. Number of students based on their ability to answer the questions

Question Number	Value Range		
	High (80–100%)	Moderate (50–79%)	Low (0–49%)
1	7	3	0
2	2	8	0
3	1	7	2
4	7	3	0
5	8	2	0
6	5	4	1
7	8	2	0
8	8	2	0

Based on the Table 5, the highest achievement of learning objectives is seen in questions 5, 7, and 8, with 8 students falling into the high category (80–100%), indicating a strong understanding. Conversely, question 3 has 2 students in the low category (0–49%), suggesting difficulties in analysis. The majority of students are in the moderate category (50–79%), especially in questions 2 and 3, indicating a sufficient understanding but with some shortcomings in analysis. Overall, the achievement of learning objectives is

fairly good, but there is a need to strengthen concepts in questions requiring deeper analysis.

After the trial of the questions, an intriguing finding was that the students wrote very clear answers and provided supporting evidence. The development of linear motion assessments using the ECD approach helps students focus on how to present evidence when answering, making their responses more directed and easier to understand. These findings indicate that the ECD approach can make assessments more more valid in measuring students' understanding.

This assessment is only at the development stage and has not yet reached the testing phase for a wider scale. It is hoped that this assessment can serve as a useful reference for teachers in designing and developing more effective assessments. The assessment developed aims to provide clear and structured guidance in evaluating students' understanding while also facilitating a more in-depth evaluation of abilities and skills relevant to learning objectives. Thus, teachers can more easily assess students' learning outcomes and adjust appropriate teaching strategies.

Conclusion

This research successfully developed a straight-line motion kinematics assessment based on ECD to improve the quality of physics learning evaluation. This assessment is designed using the ADDIE approach, which includes the stages of analysis, design, and development to comprehensively measure students' conceptual understanding. The validation results show a very high level of validity with an average Aiken index score of 0.97, while the assessment trial indicates that this instrument is suitable for evaluating the achievement of learning objectives and exploring students' deep understanding. With these results, this ECD-based assessment can serve as an appropriate instrument to measure the achievement of learning objectives more accurately. In addition, this assessment can serve as a reference for teachers in developing more structured, relevant evaluations that align with 21st-century skills. This research has only reached the development stage, so further studies are needed to test the effectiveness of the assessment in various learning contexts, evaluate its implementation in the classroom, and analyze its impact on students' understanding and achievement of learning objectives.

References

- ACER; UNESCO Institute for Statistics. 2017. *Principles of Good Practice in Learning Assessment*. ACER, Melbourne.
- Aiken, L.R. 1985. Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1):131-142. <https://doi.org/10.1177/0013164485451012>
- Apriyani, N. Ariani, T. & Arini, W. 2020. Pengembangan modul fisika berbasis discovery learning pada materi fluida statis siswa kelas XI SMA Negeri 1 Lubuklinggau tahun pelajaran 2019/2020. *Silampari Jurnal Pendidikan Ilmu Fisika*, 2(1):41-54. <https://doi.org/10.31540/sjpif.v2i1>
- Arends, R.I. 2012. *Learning to Teach* (9th Edition). McGraw-Hill Education, New York
- Arifin, Z. 2016. *Evaluasi Pembelajaran: Prinsip, Teknik, dan Prosedur*. PT Remaja Rosdakarya, Bandung

- Arikunto, S. 2012. *Dasar-dasar Evaluasi Pendidikan*. Bumi Aksara, Jakarta
- Bashooir, K., & Supahar. 2018. Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM. *Jurnal Penelitian dan Evaluasi Pendidikan*, 22(2):219-230. <http://dx.doi.org/10.21831/pep.v22i2.19590>
- Bechard, E., Clark, A., Romine, R.S., Karvonen, M., Neal, K., Erickson, K. 2019. Use of evidence-centered design to develop learning maps-based assessments. *International Journal of Testing*, 19(2):188-205. <http://dx.doi.org/10.1080/15305058.2018.1543310>
- Bhakti, Y.B., Arthur, R., & Supriyati, Y. 2023. Development of an assessment instrument for critical thinking skills in physics: a systematic literature review. *Journal of Physics: Conference Series*, 2596(1):012067. <http://dx.doi.org/10.1088/1742-6596/2596/1/012067>
- Branch, RM. 2009. *'Instructional Design: The ADDIE Approach'*. Springer, United States of America.
- Brookhart, S.M. 2010. *How to Assess Higher-Order Thinking Skills in Your Classroom*. VA: ASCD, Alexandria.
- Brookhart, S.M. 2013. *How to Create and Use Rubrics for Formative Assessment and Grading*. ASCD, Alexandria.
- Crawford, B.A., & Capps, D. K. 2018. Teacher Cognition of Engaging Children in Scientific Practices. In Y. Dori, Z. Mevarech, & D.R. Baker (Eds.), *Cognition, metacognition, and culture in STEM education: Learning, teaching and assessment* (Vol. 24, Ser. Cognition, Metacognition, and Culture in STEM Education. Innovations in Science Education and Technology). Springer International Publishing AG. 24. https://doi.org/10.1007/978-3-319-66659-4_2
- Duschl, R.A. & Bybee, R.W. 2014. Planning and carrying out investigations: An entry to learning and to teacher professional development around NGSS science and engineering practices. *International Journal of STEM Education*, 1(12):1-9. <https://doi.org/10.1186/s40594-014-0012-6>
- Ellianawati, E., Mufiatunnikmah, S., Setyaningsih, N., & Subali, B. 2020. Asesmen multi representasi berbasis keterampilan abad ke-21 pada materi gerak lurus. *Physics Education Research Journal*, 2(1):19-34. <https://doi.org/10.21580/perj.2020.2.1.5038>
- Giancoli, D. C. 2001. *Fisika (5th ed.)*. Erlangga, Jakarta.
- Harris, C. J., Krajcik, S. C., Pellegrino, J. W., & DeBarger, A. H. 2019. Designing knowledge-in-use assessments to promote deeper learning. *Educational Measurement: Issues and Practice*, 38(2):53-67. <https://doi.org/10.1111/emip.12253>
- Henderson, R. 2024. *Redesigning Legacy Conceptual Inventories: Using Evidence-Centered Design to Develop Valid, Equitable, and Flexible Items for Conceptual Understanding*. PERC 2024.

- Jufriadi, A., Sutopo., Kusairi, S., & Sunaryono. 2023. Assessment of kinematic concepts comprehension: A systematic review. *International Journal of Evaluation and Research in Education*, 12(3):1449-1459. <http://doi.org/10.11591/ijere.v12i3.24546>
- Kemendikbud. 2017. Panduan Penilaian oleh Pendidik dan Satuan Pendidikan pada Sekolah Dasar dan Sekolah Menengah Pertama. Direktorat Pembinaan Sekolah Dasar, Jakarta.
- Kemendikburistek, 2022. *Pembelajaran dan Asesmen: Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Menengah*. Badan Standar, Kurikulum, dan Asesmen Pendidikan, Jakarta
- Klein, P., Muller, A., & Kuhn, J. 2017. Assessment of representational competence in kinematics. *Physical Review Physics Education Research*, 13(1):010132, <https://doi.org/10.1103/PhysRevPhysEducRes.13.010132>
- Lager, A., & Lavonen, J. 2023. Engaging Students in Scientific Practices in a Remote Setting. *Education Sciences*: 13(5):431-438. <https://doi.org/10.3390/educsci13050431>
- Laverty, J.T., Underwood, S.M., Matz, R.L., Posey, L.A., Carmel, J.H., Caballero, M.D., Fata-Hartley, C.L., Ebert-May, D., Jardeleza, S.E., & Cooper, M.M. 2016. Characterizing college science assessments: the three-dimensional learning assessment protocol. *PLoS One*, 11(9):e0162333. <https://doi.org/10.1371/journal.pone.0162333>
- Mislevy, R.J. & Haertel, G.D. 2007. Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, 25(4):6-20. <https://doi.org/10.1111/j.1745-3992.2006.00075.x>
- Mislevy, R.J., Behrens, J.T., Dicerbo, K.E. & Levy, R. 2012. Design and discovery in educational assessment: evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining*, 4(1):11-48. <https://doi.org/10.5281/zenodo.3554641>
- Mislevy, R.J., Steinberg, L.S., & Almond, R.G. 2003. Focus article: on the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1):3-62. https://doi.org/10.1207/S15366359MEA0101_02
- Müftüoğlu, C.T., & Sahin, M. 2024. Framework of Assessment Design Based on Evidence-Centered Design for Assessment Analytics. In: Sahin, M., Ifenthaler, D. (eds) *Assessment Analytics in Education. Advances in Analytics for Learning and Teaching*. Springer, Cham. https://doi.org/10.1007/978-3-031-56365-2_8
- Newton, S., Alemdar, M., Rutstein, D., Edwards, D., Helms, M., Hernandez, D., & Usselman, M. 2021. Utilizing Evidence-Centered Design to Develop Assessments: A High School Introductory Computer Science Course. *Frontiers in Education*, 6(June), 1-19.
- Pellegrino, J.W., DiBello, L.V., & Goldman, S.R. 2016. A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist*, 51(1):59-81. <http://dx.doi.org/10.1080/00461520.2016.1145550>

- Pellegrino, J.W., Wilson, M.R., Koenig, J.A., & Beatty, A.S. 2014. *Developing assessments for the Next Generation Science Standards*. National Academies Press, USA.
- Poutot, G., & Blandin, B. 2015. Exploration of students' misconceptions in mechanics using the FCI. *American Journal of Educational Research*, 3(2):116-120. <http://dx.doi.org/10.12691/education-3-2-2>
- Rahmadani, U., Sundari, P.D., Hidayati, H., & Dewi, W.S. 2023. Tinjauan sistematis kesalahpahaman dalam kinematika: identifikasi, penyebab, dan remediasi. *Jurnal Penelitian Pendidikan IPA*, 9(12):1274-1283. <https://doi.org/10.29303/jppipa.v9i12.4956>
- Santrock, J.W. 2011. *Educational Psychology*. McGraw-Hill Education, Amerika Serikat.
- Saribas, D., & Ceyhan, G.D. 2015. Learning to teach scientific practice: pedagogical decisions and reflections during a course for pre-service science teacher. *International Journal of STEM Education*, (2)7:1-13. <https://doi.org/10.1186/s40594-015-0023-y>
- Sirait, J. 2024. Merancang asesmen fisika menggunakan evidence-centered design (ECD) dan three-dimensional learning protocol (3D-LAP). *Jurnal Pendidikan Informatika dan Sains*, 13(2):165-175. <https://doi.org/10.31571/saintek.v13i2.8228>
- Sirait, J. & Oktavianti, E. 2021. Pengembangan dan validasi angket motivasi belajar fisika (AMBF): Studi pilot. *Jurnal Ilmiah Pendidikan Fisika*, 5(3):305-316. <https://doi.org/10.20527/jipf.v5i3.3829>
- Sirait, J., Firdaus, Hidayatullah, M.M.S., & Habellia, R.C. 2023. Development and validation of force test to assess physics education students' representational competence. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, 11(1):306-317. <https://dx.doi.org/10.24815/jpsi.v11i2.28178>
- Sirait, J., Ricky, P., & Hamdani. 2023. Development and validation of the force diagrams representation test. *JPPPF (Jurnal Penelitian dan Pengembangan Pendidikan Fisika)*. 9(2):183-196. <https://doi.org/10.21009/1.09201>
- Steinkuehler, C. 2011. *Learning Science Through Computer Games and Simulations*. National Academies Press, USA.
- Stoen, S.M., McDaniel, M.A., Frey, R.F., Hynes, K.M., & Cahill, M.J. 2020. Force Concept Inventory: More than just conceptual understanding. *Physical Review Physics Education Research*, 16(1):010105. <https://doi.org/10.1103/PhysRevPhysEducRes.16.010105>
- Sudijono, A. 2016. *Pengantar Evaluasi Pendidikan*. Rajawali Pers, Jakarta.
- Sugiyono. 2017. *Metodologi Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta, Bandung
- Weisdiyanti, N., & Juliani, R. 2023. Development of critical thinking skills-based physics test instruments for High School. *Proceedings of the 4th International Conference on Science and Technology Applications, ICoSTA 2022, 2022 November, Medan, North Sumatera Province, Indonesia*.

- Widodo, K. Handoyo, E., & Masyhar, A. 2021. Performance of educational assessments: integrated assessment as an assessment innovation during the covid-19 pandemic. *Turkish Journal of Computer and Mathematics Education*, 12(6):2708-2718. <https://doi.org/10.17762/turcomat.v12i6.5777>
- Windschitl, M., Thompson, J., & Braaten, M. 2018. *Ambitious Science Teaching*. Harvard Education Press, London
- Yuliantaningrum, L. & Sunarti, T. 2020. Pengembangan instrumen soal HOTS untuk mengukur keterampilan berpikir kritis, berpikir kreatif, dan pemecahan masalah materi gerak lurus pada peserta didik SMA. *IPF: Inovasi Pendidikan Fisika*, 9(2):75-82. <http://dx.doi.org/10.30872/diglosia.v5i2.414>