https://research.adra.ac.id /index.php/ijen/
P - ISSN: 2988-1579
E - ISSN: 2988-0092

# Making Physics Reasoning Visible: A Narrative-Informed Design of Embedded Assessment in Collaborative Learning

**Imam Permana[1], Misykat Malik Ibrahim[2], St. Syamsudduha[3], Ulfiani Rahman[4], Fitriani Kadir[5]**

[1]Universitas Islam Negeri Alauddin, Indonesia
[2]Universitas Islam Negeri Alauddin, Indonesia
[3]Universitas Islam Negeri Alauddin, Indonesia
[4]Universitas Islam Negeri Alauddin, Indonesia
[5]Universitas Muslim Maros, Indonesia

## ABSTRACT

**Background.** Collaborative physics learning can make classroom participation visible, yet the reasoning behind students' answers often remains difficult to follow. This concern is especially relevant in teacher education, where pre-service physics teachers are expected to understand physical concepts and learn how explanations are formed, challenged, revised, and consolidated.

**Purpose.** This study examined how embedded assessment within a collaborative digital learning trajectory made students' reasoning visible during instruction on Simple Harmonic Motion.

**Method.** A narrative-informed mixed-methods design account was used. Qualitative evidence was collected from digital worksheet responses, classroom observations, and student reflections, while quantitative learning outcome data served as supporting evidence. The learning sequence was organized into three episodes—ignition, cruise, and terminal—and was supported by a microsite and interactive digital worksheets. The analysis focused on how students' written explanations, revisions, and reflections changed across these episodes.

**Results.** The findings show that embedded assessment documented students' initial intuitive ideas, captured conceptual negotiation during collaborative work, and preserved reflective consolidation after instruction. Students' explanations gradually shifted from descriptive accounts of repeated motion toward more relational reasoning involving displacement, restoring force, velocity, acceleration, and equilibrium. Quantitative evidence also indicated improvement in conceptual understanding after initial ability was controlled, supporting the patterns found in the process data.

**Conclusion**. Embedded assessment can function as narrative evidence of reasoning development when it is deliberately connected to collaborative tasks, digital documentation, and a coherent learning trajectory. This study offers a practical framework for using digital worksheets not only as task spaces, but also as tools for tracing, interpreting, and supporting reasoning development in collaborative physics teacher education.

## KEYWORDS

Collaborative Learning; Digital Physics Learning; Embedded Assessment; Learning Trajectory; Student Reasoning

**Citation:** Permana, P., Ibrahim, M. M., Syamsudduha, S., Rahman, U., & Kadir, F. (2026). Making Physics Reasoning Visible: A Narrative-Informed Design of Embedded Assessment in Collaborative Learning. *International Journal of Educational Narrative*, *4*(3), 783–799. https://doi.org/10.70177/ijen.v4i3.3856

**Correspondence:**

Imam Permana,
Imam.permana922@gmail.com.

**Received:** January 19, 2026

**Accepted:** March 15, 2026

**Published:** June 25, 2026



## INTRODUCTION

Collaborative learning has become an important pedagogical approach in physics education because it provides opportunities for students to exchange ideas, negotiate meanings, and construct understanding through social interaction (Barkley et al., 2014; W. Johnson & T. Johnson, 2019) However, participation in collaborative tasks does not automatically indicate that students' reasoning is visible. Students may appear active in discussion, complete group worksheets, or arrive at correct answers while still relying on procedural strategies that do not reveal how they connect physical concepts. This problem is particularly significant in physics learning, where conceptual understanding depends not only on obtaining correct numerical results but also on explaining why a phenomenon occurs and how one physical quantity relates to another. For pre-service physics teachers, this issue becomes even more important. They are not only learners of physics content but also future teachers who must be able to interpret students' thinking, identify conceptual difficulties, and design appropriate instructional responses. Therefore, collaborative physics learning needs to be designed not merely to increase classroom participation but to make students' reasoning traceable during the learning process.

Simple harmonic motion (SHM) is a relevant context for examining this issue because it requires students to understand a system of interrelated physical quantities. Learning SHM is not limited to recognizing periodic motion or using equations for period, frequency, and spring constants. Students must coordinate displacement, restoring force, velocity, acceleration, equilibrium position, and energy transformation within a dynamic system (Ainsworth, 2008; Nugraha et al., 2020). Previous studies have shown that students often experience difficulty in constructing these relationships, even when they can solve routine mathematical problems (Somroob & Wattanakasiwich, 2017). For instance, students may write $F = -kx$ correctly but remain unable to explain why the restoring force always acts opposite to displacement. In pendulum motion, they may also overgeneralize the role of mass or confuse the relationship between speed, force, and acceleration at different positions of the swing. These difficulties show that the core challenge in SHM is not only computational accuracy but the development of coherent relational reasoning.

The persistence of fragmented reasoning in SHM has direct implications for physics teacher education. Pre-service teachers who understand SHM only as a set of formulas may later reproduce formula-centered instruction in their own classrooms. They may be able to solve textbook problems but remain less prepared to explain conceptual relations, anticipate misconceptions, or guide students from everyday intuition toward scientific explanation. Research on teacher education emphasizes that prospective teachers need opportunities to develop both content knowledge and the ability to notice, interpret, and respond to students' thinking (Jacobsson et al., 2024; Schiering et al., 2023). Thus, making reasoning visible is not only a strategy for improving students' conceptual understanding. It is also a way of preparing pre-service teachers to recognize how physics concepts are constructed, challenged, and reorganized through learning.

One reason students' reasoning often remains hidden is that assessment is frequently positioned as an endpoint of instruction. Pretests, posttests, and summative achievement measures are useful for identifying whether learning outcomes improve, but they provide limited information about how students' reasoning develops during instruction (Black & Wiliam, 2018; William & Thompson, 2017; McMillan, 2024). A higher posttest score may indicate improvement, yet it does not necessarily show which ideas students initially held, which conceptual conflicts emerged, which explanations were revised, or which representations supported their understanding. In other words, conventional assessment can show evidence of learning achievement but may fail to capture the pathway of reasoning that led to that achievement. What remains underexamined is not simply whether students improve, but how their reasoning becomes traceable while learning is taking place. This distinction is essential in physics teacher education because learning to reason conceptually is closely related to learning how to teach conceptually.

Embedded assessment offers a way to address this limitation by integrating assessment into the instructional process rather than placing it only before or after learning activities. When assessment is embedded in learning, students' predictions, written explanations, diagrams, justifications, revisions, and reflections can become evidence of developing understanding (Black & Wiliam, 2018; Clark & Karvonen, 2021). This evidence becomes more pedagogically meaningful when assessment tasks are authentic, cognitively demanding, and connected to meaningful learning contexts rather than limited to final-answer checking (Villarroel et al., 2018). In this study, embedded assessment is integrated with Authentic Assessment Based on Teaching Learning Trajectory (AABTLT), which organizes learning according to anticipated stages of students' conceptual development (Rochman et al., 2017). The learning design consists of three instructional episodes: ignition, cruise, and terminal. The ignition episode elicits students' initial ideas and everyday experiences related to oscillatory motion. The cruise episode facilitates conceptual negotiation through collaborative tasks, digital activities, simulations, and peer explanation. The terminal episode guides students toward consolidation and reflection. Through this structure, assessment is not separated from instruction but becomes part of the trajectory through which reasoning is documented.

The digital environment used in this study strengthens the documentation function of embedded assessment. A microsite was used to organize learning resources, while interactive digital worksheets provided spaces for students to respond to prompts, construct explanations, and submit learning artifacts. Digital worksheets can preserve students' written responses and representations in ways that ordinary oral discussion often cannot (Munfaridah et al., 2021). However, the pedagogical value of digital worksheets does not lie in the technology itself. Their value depends on how the prompts are designed and how they invite students to explain, compare, justify, revise, and reflect. When aligned with a learning trajectory, digital worksheets can help document how students move from intuitive reasoning toward more relational conceptual explanations. In this sense, the digital platform functions not only as a medium for delivering tasks but also as a space for recording the development of students' reasoning.

Several strands of previous research provide important foundations for this study. Studies on collaborative learning have emphasized the role of peer interaction, shared regulation, and collective knowledge construction in supporting active learning (Barkley et al., 2014; Ucan & Webb, 2015). Research on formative, authentic, and embedded assessment has shown that assessment can support learning when it provides evidence for feedback, reflection, and instructional decision-making (Black & Wiliam, 2018; Clark & Karvonen, 2021; William & Thompson, 2017). Studies on learning trajectories have highlighted the importance of sequencing learning tasks according to anticipated conceptual development (Ferreira & Lemmer, 2021; Rochman et al., 2017). Research on digital learning environments has also indicated the potential of digital tools to support interaction, access, and documentation (Munfaridah et al., 2021). Nevertheless, these areas are often treated separately. Collaboration is commonly discussed as a learning strategy, assessment as a feedback or measurement practice, learning trajectory as instructional sequencing, and digital worksheets as technological tools. Less attention has been given to how these elements can be integrated to make students' reasoning visible across a collaborative physics learning trajectory.

This gap is both methodological and conceptual. If collaborative learning is treated only as group work, digital worksheets only as media, assessment only as measurement, and learning trajectory only as task sequencing, then the development of students' reasoning may remain analytically fragmented. A more integrated perspective is needed to explain how students' ideas emerge, encounter conceptual tension, become negotiated through interaction, and are reorganized into more coherent explanations. This study responds to that need by positioning embedded assessment as a mechanism for generating narrative evidence of reasoning development. The term narrative evidence does not refer to anecdotal storytelling. Rather, it refers to learning evidence that can be interpreted as a connected sequence of conceptual movement across instructional episodes.

To support this interpretation, this study adopts a narrative-informed lens. Narrative has been widely understood as a way of organizing experience, meaning, and development across time (Bruner, 1990; Clandinin & Connelly, 2000). In science education, narrative-oriented perspectives can help researchers examine how learners connect events, construct explanatory coherence, and make sense of conceptual change as a process rather than as a single outcome (T. O. Odden, 2021). In this study, narrative is used as an analytical lens for interpreting students' written responses, collaborative explanations, and reflective statements as a sequence of reasoning development. The ignition, cruise, and terminal episodes are therefore not only instructional phases but also analytical episodes. Ignition reveals students' initial reasoning, cruise reveals negotiation and reorganization, and terminal reveals consolidation and reflection. This perspective allows the study to move beyond isolated worksheet answers or test scores toward a developmental account of learning.

The main novelty of this study lies in positioning embedded assessment within a digital collaborative learning environment as narrative evidence of students' reasoning development. This proposition contributes to a shift from viewing assessment as a static audit of achievement toward understanding it as a dynamic record of conceptual movement. The study does not assume that every digital response represents deep learning, nor does it claim that collaboration automatically produces conceptual change. Instead, it argues that when assessment prompts are embedded within a learning trajectory and connected to collaborative activity, they can generate evidence for tracing how students' explanations develop. Conceptually, this study extends discussions of assessment by emphasizing its narrative function. Methodologically, it integrates digital worksheet responses, classroom observations, reflective statements, and supporting learning outcome evidence. Practically, it offers physics educators a model for designing worksheets as tools for tracking reasoning rather than merely instruments for completing tasks.

Consequently, this study examines how embedded assessment within a collaborative digital physics learning design makes students' reasoning visible during instruction on Simple Harmonic Motion. The inquiry is guided by the following research questions:

1. How was embedded assessment designed within a collaborative digital learning trajectory for teaching Simple Harmonic Motion?
2. How did students' reasoning patterns shift across the ignition, cruise, and terminal episodes?
3. How did learning outcome evidence support the interpretation of students' reasoning development?

To answer these questions, the study employs a narrative-informed mixed-methods design account. Qualitative process data were drawn from students' digital worksheet responses, classroom observations, and reflective statements, while quantitative learning outcome evidence was used to support the interpretation of conceptual development. Through this approach, the study seeks to show how collaborative physics learning can be designed not only to engage students in visible activity but also to make the development of their reasoning pedagogically visible.

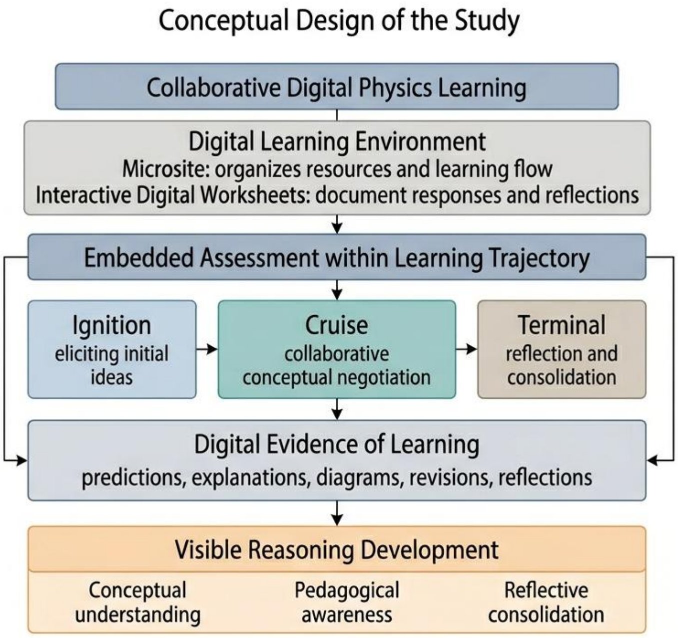


**Figure 1.** Conceptual design for documenting students' reasoning development.

## RESEARCH METHODOLOGY

### Research Design

This study used a narrative-informed mixed-methods design account to examine how embedded assessment in a collaborative digital physics learning environment made students' reasoning visible during instruction. The mixed-methods orientation was selected because the study required both process evidence and outcome evidence to understand learning more completely (Creswell & Creswell, 2023; Creswell & Plano Clark, 2018). The qualitative strand was treated as the central source of evidence because the study was primarily concerned with how students' explanations developed across learning episodes. Digital worksheet responses, classroom observations, and reflective statements were analyzed as traces of reasoning development. Quantitative learning outcome data were used as a supporting strand to examine whether the reasoning shifts identified in the process data were accompanied by measurable improvement in conceptual understanding. The narrative-informed orientation allowed students' responses to be interpreted not as isolated products, but as connected episodes in the development of physics reasoning (Clandinin & Connelly, 2000; T. O. B. Odden & Russ, 2019).

### Population and Sample

The study was conducted in a Basic Physics course in a Physics Education program at a public university in Indonesia. The selected topic was Simple Harmonic Motion (SHM), because this topic requires students to coordinate displacement, restoring force, velocity, acceleration, and energy transformation within a dynamic system. The participants were pre-service physics teachers enrolled in two intact classes that were scheduled to study SHM during the implementation period. Because the research was situated in an authentic classroom setting, random assignment was not applied. The use of intact classes is common in quasi-experimental educational research, particularly when instructional schedules and class structures cannot be reorganized for research purposes (Shadish et al., 2002). A purposive-convenience sampling strategy was used to select classes that matched the instructional topic, timetable, and digital access requirements. One class became the main site for the narrative analysis because it received the collaborative digital learning design with embedded assessment, while the comparison class was used to support the interpretation of quantitative learning outcome evidence.

**Learning Design and Digital Environment**

The learning environment developed for this study was an integrated digital ecosystem called Fiskonautra 3C, referring to concept, competency, and character. It consisted of a central microsite connected to interactive digital worksheets. The microsite organized learning materials, instructions, collaborative tasks, and supporting learning features. The digital worksheets provided spaces for students to write explanations, draw representations, make predictions, justify answers, revise ideas, and write reflections. The instructional design was guided by Authentic Assessment Based on Teaching Learning Trajectory (AABTLT), which places assessment within the learning process and organizes instruction according to anticipated stages of students' conceptual development (Rochman et al., 2018). This approach is also consistent with the broader logic of embedded and formative assessment, where assessment evidence is used to understand and support learning while instruction is still unfolding (Black & Wiliam, 2018; Clark & Karvonen, 2021). In this study, the learning trajectory was organized into three episodes: ignition, cruise, and terminal. Ignition elicited students' initial ideas and everyday experiences of oscillatory motion. Cruise supported collaborative conceptual negotiation through tasks, simulation-based activities, and peer explanation. Terminal guided students toward consolidation and reflection.

**Research Instruments and Data Sources**

Data were collected from four sources. First, digital worksheet responses became the main evidence of reasoning development because they documented how students explained, represented, revised, and reflected on SHM concepts across the instructional episodes. Second, classroom observations were used to document the implementation of collaborative learning, students' use of the digital worksheets, and the interaction patterns that emerged during learning. Observation was important because collaborative learning should not be reduced to group formation; it needs to be examined through the quality of explanation, negotiation, and shared construction of meaning among learners (Barkley et al., 2014; Ucan & Webb, 2015). Third, student reflections were collected to capture how students consolidated their understanding at the end of the learning sequence. Fourth, a conceptual understanding test was used to provide supporting quantitative evidence of learning outcome change. The test consisted of multiple-choice items on SHM and had been reviewed by expert validators before implementation. Reliability analysis showed high internal consistency, with a Kuder-Richardson 20 coefficient of 0.938. KR-20 was used because the test items were scored dichotomously as correct or incorrect (DeVellis, 2017). Table 1 presents the data sources used in the study and clarifies how each source contributed to the analysis of students' reasoning development.

**Table 1.** Data Sources and Analytic Roles

| Data Source | Function in the Study | Analytic Role |
|---|---|---|
| Digital worksheet responses | Captured written explanations, diagrams, revisions, and reflections across learning episodes | Main evidence of reasoning development |
| Classroom observations | Documented collaborative interaction, digital worksheet use, and implementation fidelity | Contextual evidence |
| Student reflections | Captured consolidation of conceptual and reflective understanding | Narrative evidence |
| Conceptual understanding test | Measured learning outcome change before and after instruction | Supporting quantitative evidence |

## Procedures

The research procedure followed the structure of the teaching learning trajectory. Before instruction, students completed a conceptual understanding pretest to identify their initial understanding of SHM. During the instructional sequence, students accessed materials through the microsite and completed embedded assessment prompts in the digital worksheets. In the ignition episode, they responded to prompts designed to elicit initial reasoning about oscillatory motion. In the cruise episode, they worked collaboratively on conceptual tasks and simulation-supported activities to negotiate explanations with peers. In the terminal episode, they completed consolidation and reflection prompts. Classroom observers documented how the learning design was implemented and how students interacted during collaborative activities. After the instructional sequence, students completed a posttest, and their digital worksheet responses were exported as stored learning artifacts for analysis. Each instructional session lasted approximately 150 minutes, following the regular course schedule. Figure 2 summarizes the procedural flow of the study by linking the instructional episodes, data collection points, and integration of qualitative and quantitative evidence.

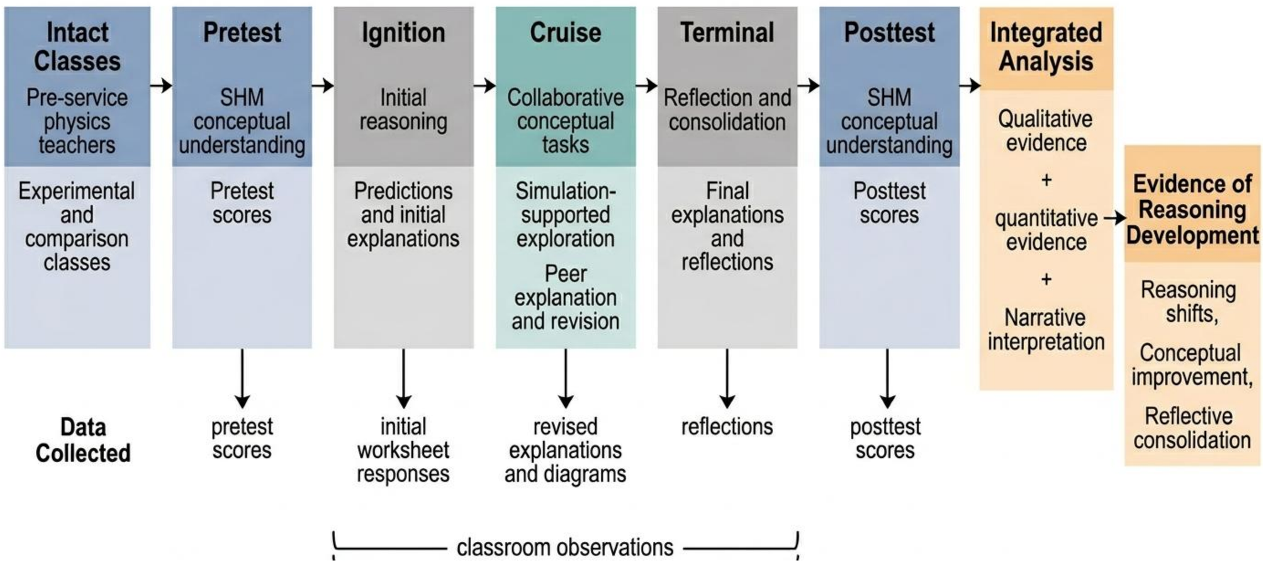


**Figure 2.** Research procedure across instructional episodes.

## Data Analysis

Qualitative analysis focused on interpreting students' responses as evidence of reasoning development. The unit of analysis included written explanations, diagrams, justifications, revisions, and reflective statements. Students' responses were compared across the ignition, cruise, and terminal episodes to identify whether their initial ideas were maintained, revised, or reorganized during instruction. The analysis did not merely count correct answers; it focused on changes in the quality of reasoning. Initial analytic categories included intuitive reasoning, descriptive reasoning, relational reasoning, representational coordination, collaborative revision, and reflective consolidation. These categories were refined through repeated reading of worksheet responses and observation notes. Qualitative coding and analytic memoing were used to organize patterns, compare evidence across data sources, and develop interpretations grounded in students' actual responses (Miles et al., 2014; Saldana, 2021). Observation notes helped contextualize the worksheet data by showing how collaborative interaction supported or limited the development of students' explanations.

Quantitative analysis was used as a supporting strand. Pretest and posttest scores were analyzed to examine changes in students' conceptual understanding. When comparing the experimental and comparison classes, Analysis of Covariance (ANCOVA) was used, with pretest scores treated as the covariate. ANCOVA was selected because it allows posttest comparisons while adjusting for initial differences between groups, a procedure commonly used in quasi-experimental educational studies when assumptions are examined properly (Field, 2024; Tabachnick & Fidell, 2019). Before the main analysis, assumptions of normality, homogeneity of variance, and homogeneity of regression slopes were tested. Partial eta squared was used to interpret effect size because statistical significance alone does not indicate the practical magnitude of an instructional effect (Cohen, 1988; Ferguson, 2009). When an assumption violation was identified, bootstrap ANCOVA was used as a robustness check to examine whether the inference remained stable under resampling (Efron & Tibshirani, 1993; Field, 2024). The quantitative results were interpreted as supporting evidence, not as a replacement for the narrative analysis of reasoning development.

**Validity, Reliability, and Trustworthiness**

Several strategies were applied to strengthen the rigor of the study. Instrument validity was examined through expert judgment and content validity analysis. The Content Validity Index was used to evaluate the degree of expert agreement regarding item relevance, clarity, and representativeness (Lynn, 1986; Polit & Beck, 2006; Yusoff, 2019). Reliability was examined using KR-20 for the conceptual understanding test. Trustworthiness of the qualitative interpretation was supported through triangulation across digital worksheet responses, classroom observations, and student reflections. Triangulation was used to reduce dependence on a single source of evidence and to strengthen the credibility of interpretation (Lincoln & Guba, 1985; Miles et al., 2014) Interpretive claims were checked against the sequence of learning activities so that conclusions about reasoning development remained grounded in students' responses. An audit trail was maintained by organizing digital artifacts, observation notes, coding decisions, and analytic memos. The study has several limitations. It used intact classes rather than random assignment, was conducted in one institutional context, and focused on a limited instructional cycle on SHM. Therefore, the study does not claim broad statistical generalization. Instead, it offers an analytically grounded account of how embedded assessment can make students' reasoning visible within a specific collaborative digital physics learning context.

## RESULTS AND DISCUSSION

The findings show that the collaborative digital learning design did more than organize students' activities. It created a documented pathway through which students' reasoning could be followed from their initial ideas, through collaborative negotiation, toward later consolidation. This visibility emerged because assessment was not placed only at the end of instruction. Instead, assessment prompts were embedded in the learning sequence and required students to externalize their predictions, explanations, diagrams, revisions, and reflections. As a result, students' written responses became learning artifacts that preserved how their understanding of simple harmonic motion developed across the lesson.

The first important finding concerns the way embedded assessment worked across the learning trajectory. The ignition episode elicited students' starting points. At this stage, students tended to describe oscillatory motion through everyday experience: a swing moving back and forth, an object returning repeatedly, or a motion that feels faster or slower depending on visible changes in the system. These responses were valuable not because they were fully scientific, but because they revealed the intuitive resources students brought into the lesson. The cruise episode then used collaborative tasks, simulation-supported exploration, and worksheet prompts to make students compare their initial assumptions with observed patterns. The terminal episode asked students to synthesize and reflect, allowing the instructor to see whether students could reconnect the ideas discussed earlier into a more coherent explanation.

**Table 2.** Embedded Assessment Across the Learning Trajectory

| Episode | Assessment Prompt Function | Evidence Captured | Interpretation |
|---|---|---|---|
| Ignition | Eliciting prior ideas and experiential reasoning | Predictions, everyday explanations, initial diagrams | Reveals intuitive or descriptive reasoning |
| Cruise | Supporting comparison, justification, and conceptual negotiation | Group explanations, revised answers, simulation-based representations | Shows transitional reasoning and conceptual reorganization |
| Terminal | Consolidating conceptual understanding and reflection | Final explanations, reflective responses, conceptual summaries | Indicates reflective consolidation and movement toward relational reasoning |

Table 2 indicates that each episode generated a different layer of evidence. The ignition episode exposed how students initially framed the phenomenon. The cruise episode documented how these ideas were tested and revised through collaboration. The terminal episode showed whether students could produce a more organized explanation after the learning process. In this study, the digital worksheet became the medium through which students' responses, revisions, and reflections were documented across the instructional sequence.

The second finding concerns the form of reasoning development observed in students' responses. At the beginning, many explanations were descriptive. Students could identify the visible feature of oscillation, but they did not yet explain the mechanism that sustains it. For example, oscillation was often described as repeated motion without clear reference to restoring force, equilibrium, or the changing relationship among displacement, velocity, and acceleration. This indicates that students could recognize the phenomenon but still struggled to explain why the system behaved as it did.

During the cruise episode, students' reasoning began to move toward more relational accounts. The shift was visible when students started linking the direction of restoring force to displacement from equilibrium, or when they began to distinguish between position, velocity, and acceleration at different points of motion. These explanations were not always complete. Some students still used partially fragmented reasoning, especially when moving from qualitative description to a more systematic account of the motion. Nevertheless, the responses showed a meaningful movement: SHM was no longer treated only as "motion that repeats," but increasingly as a system of connected physical quantities.

The terminal episode provided a more cautious picture of consolidation. Some students were able to connect equilibrium position, maximum displacement, restoring force direction, and changing velocity in a more coherent explanation. Others could name the relevant variables but still had difficulty articulating their relationships consistently. This unevenness is important because it showed which aspects of reasoning had developed and which remained unstable. Figure 3 synthesizes the progression of students' reasoning across the instructional episodes, from intuitive descriptions toward more relational, although uneven, explanations.

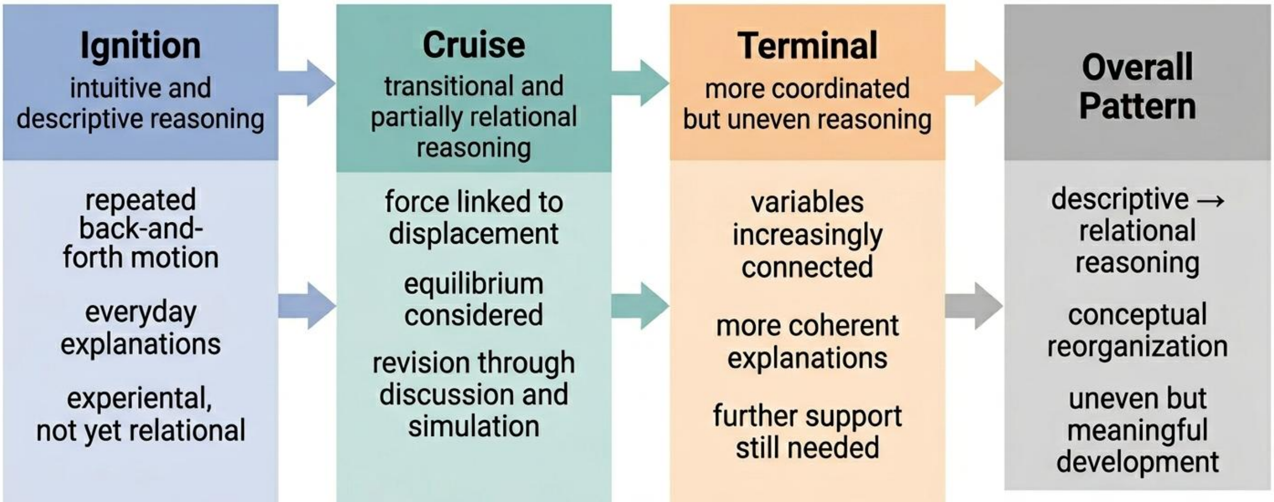


**Figure 3**. Development of students' reasoning across instructional episodes.

The third finding is that collaboration supported reasoning development only when it was structured by prompts that demanded explanation. In the cruise episode, students were not merely asked to complete a group task. They were asked to compare predictions, justify answers, discuss simulation results, and revise explanations. This structure encouraged students to expose their thinking to peers and to confront inconsistencies between intuitive claims and observed evidence.

However, the quality of collaboration was not uniform. Some groups showed genuine exchange of reasoning, while others moved quickly toward agreement without sustained conceptual discussion. This variation shows that collaborative learning differed across groups. The embedded assessment prompts helped reduce superficial collaboration because students had to leave written evidence of their reasoning, but they did not remove all implementation constraints. Time management and clarity of digital instructions remained important issues during classroom use.

The terminal reflections added another layer to the findings. For pre-service physics teachers, reflection did not only function as a summary of the lesson. It also created a space for students to begin thinking about the teachability of the concept. Some responses suggested early pedagogical awareness, especially when students recognized that learners might misunderstand the role of mass in pendulum motion, the direction of restoring force, or the difference between speed and acceleration. The evidence indicates that students began to shift from learning Simple Harmonic Motion only as physics content toward noticing that the same content may become difficult for future learners.

The quantitative evidence was used to support the qualitative patterns identified in students' reasoning development. Before interpreting the ANCOVA result, the required assumptions were examined. The Shapiro–Wilk test showed that the pretest scores were normally distributed in both the comparison class, $p = 0.061$, and the experimental class, $p = 0.646$. For the posttest scores, the experimental class also met the normality assumption, $p = 0.740$, whereas the comparison class showed a departure from normality, $p = 0.004$. Levene's test indicated that the homogeneity of variance assumption was satisfied, $p = 0.231$. The homogeneity of regression slopes assumption was also met because the interaction between group and pretest score was not significant, $p = 0.807$. These results indicated that the main ANCOVA assumptions were generally acceptable, although the non-normal posttest distribution in the comparison class required additional caution. Therefore, bootstrap analysis was used as a robustness check.

After controlling for pretest scores, ANCOVA showed a significant difference in conceptual understanding between the experimental and comparison classes, $F(1, 41) = 5.121$, $p = 0.029$, with a moderate effect size, $\eta p^2 = 0.111$. The bootstrap ANCOVA with 5,000 resamples confirmed the stability of this result, B = 0.859, p = 0.026, with a BCa 95% confidence interval of 0.145–1.608. Because the confidence interval did not include zero, the inference was considered stable. These results suggest that the collaborative digital learning design with embedded assessment was associated with stronger conceptual understanding after students' initial ability was statistically controlled. Table 3 summarizes the assumption checks and quantitative learning outcome evidence.

**Table 3.** Summary of ANCOVA Assumption Checks and Learning Outcome Evidence

| Analysis | Result | Interpretation |
| --- | --- | --- |
| Pretest normality, comparison class | Shapiro–Wilk p = 0.061 | Normality assumption met |
| Pretest normality, experimental class | Shapiro–Wilk p = 0.646 | Normality assumption met |
| Posttest normality, comparison class | Shapiro–Wilk p = 0.004 | Normality issue identified |
| Posttest normality, experimental class | Shapiro–Wilk p = 0.740 | Normality assumption met |
| Homogeneity of variance | Levene p = 0.231 | Assumption met |
| Homogeneity of regression slopes | Group × Pretest p = 0.807 | Assumption met |
| ANCOVA | $F(1, 41) = 5.121$, p = 0.029 | Significant difference after controlling for pretest |
| Effect size | $\eta p^2 = 0.111$ | Moderate practical effect |
| Bootstrap ANCOVA | B = 0.859, p = 0.026 | Result remained significant under resampling |
| BCa 95% confidence interval | 0.145–1.608 | Confidence interval did not include zero |

The quantitative result should be read as supporting evidence, not as the whole explanation. It indicates that the experimental class showed stronger conceptual understanding after the intervention, but the score alone does not show how students arrived at that improvement. The worksheet responses and classroom observations provide the process account that the score cannot fully capture. They show that students were repeatedly asked to express initial reasoning, test ideas through collaboration, compare representations, and consolidate their explanations. Therefore, the strongest interpretation is not that the digital product by itself caused improvement. Rather, the integrated design learning trajectory, embedded assessment, collaborative tasks, and digital documentation was associated with both measurable conceptual gain and visible development in students' explanations.

The results are consistent with the logic of embedded assessment, where evidence of learning is gathered within instruction and used to understand students' developing competence rather than merely to certify final achievement (Clark & Karvonen, 2021; Wiliam & Thompson, 2017). They also reflect the central intention of Authentic Assessment Based on Teaching Learning Trajectory, namely that assessment should accompany the learning process and help trace how students' thinking develops through student activity sheets and instructional stages (Rochman et al., 2017, 2018). In this study, the digital worksheet operationalized this intention by preserving students' explanations across the ignition, cruise, and terminal episodes.

The observed shift from descriptive to more relational reasoning is also aligned with previous research in physics education. Students may perform calculations while still lacking a coordinated understanding of the conceptual structure behind formulas (Docktor & Mestre, 2014; Somroob & Wattanakasiwich, 2017). In Simple Harmonic Motion, this issue becomes visible when students can recognize repeated motion or use equations but still struggle to explain the relationships among displacement, restoring force, velocity, acceleration, and equilibrium. This matters because conceptual understanding in science requires students to organize relations among concepts and representations rather than simply recall separate facts or equations (Ainsworth, 2008; Holme et al., 2015). The findings of this study suggest that embedded assessment can make such partial understanding more visible, rather than concealing it behind final scores.

The findings also clarify the role of collaboration. Collaborative learning should not be reduced to placing students in groups; it must be intentionally designed so that learners participate in shared meaning-making and assume responsibility for developing ideas together (Barkley et al., 2014; Dillenbourg, 1999; Davidson & Major, 2014). In this study, collaboration became more meaningful when students were required to compare predictions, justify explanations, and revise responses. This supports studies in physics education showing that group problem solving is shaped by both conceptual reasoning and social interaction (Hoehn et al., 2020; Koerfer & Gregorcic, 2024). At the same time, the uneven quality of group interaction confirms that collaboration is not automatically productive. Its value depends on task structure, prompt clarity, teacher facilitation, and available learning time.

The results for pedagogical and reflective dimensions were more restrained. Some evidence suggested that students began to recognize possible learning difficulties related to SHM, but this awareness was not yet consistent across responses. This is theoretically reasonable. Pedagogical reasoning requires more than conceptual understanding. It requires students to anticipate learners' thinking, select representations, formulate instructional questions, and make decisions about how to respond to misconceptions. Such competence normally develops through repeated practice, feedback, and reflection over time (Desimone & Garet, 2015; Shulman, 1986; Schiering et al., 2023). Similarly, reflective and spiritual meaning-making appeared in some responses, but the depth varied. This suggests that affective and value-oriented outcomes may require longer cycles of experience and reflection than those available in a limited instructional sequence.

Several alternative explanations must also be considered. Part of the improvement in conceptual understanding may have resulted from repeated exposure to SHM tasks, increased familiarity with the test format, or the motivational effect of using a new digital learning environment. The use of intact classes also means that group characteristics and classroom dynamics may have influenced the results. These possibilities prevent an inflated interpretation of the findings. Yet the qualitative data suggest that the change was not limited to posttest performance. Students' written explanations showed movement from descriptive toward more relational accounts across the episodes. The process evidence therefore gives a more defensible interpretation: the intervention helped make reasoning visible and provided opportunities for students to refine it, although the extent of development varied among students and groups.

The theoretical contribution of the study lies in repositioning embedded assessment as narrative evidence of reasoning development. Assessment is often discussed as a tool for measuring learning or generating feedback. This study suggests a further function: when embedded within a learning trajectory, assessment can preserve the sequence through which students' explanations unfold. The trajectory provides the temporal structure, collaborative tasks provide the social space for negotiation, and the digital worksheet preserves the written evidence. This extends the discussion of assessment for learning and assessment as learning by foregrounding its role in making reasoning visible across instructional episodes (Wiliam & Thompson, 2017; McMillan, 2024).

The practical implication is direct. Physics educators should not treat digital worksheets as online versions of conventional worksheets. Their pedagogical power depends on the quality of the prompts. Prompts need to ask students to predict, explain, justify, compare, revise, and reflect. Without those demands, digital tools may only accelerate task completion without deepening reasoning. The findings also suggest that the value of a microsite or worksheet platform depends on its integration with instructional design. Technology becomes pedagogically meaningful when it helps organize learning, document student thinking, and inform instructional decisions. This view is consistent with the argument that educational technology should be understood as part of a broader pedagogical design, not as an isolated technical solution (Kim & Reeves, 2007; Laurillard, 2012).

The study has several limitations. It was conducted in one institutional context and used intact classes, so the findings should not be generalized statistically to all pre-service physics teacher programs. The instructional cycle was also limited to SHM, while other physics topics may involve different conceptual difficulties and representational demands. In addition, the data relied heavily on written responses. Some important aspects of peer negotiation may have occurred orally but were not fully captured in the digital artifacts. Future studies should examine the design across different physics topics, extend the implementation period, and include richer multimodal data such as video-recorded group discussion, screen recordings, or discourse analysis. These additions would make it possible to examine reasoning visibility not only in written products but also in the interactional processes through which students construct meaning.

Overall, the findings show that embedded assessment within a collaborative digital learning trajectory can make students' reasoning more visible by documenting initial ideas, capturing conceptual negotiation, and preserving reflective consolidation. The strongest evidence appeared in conceptual understanding, where qualitative shifts in explanations were supported by significant quantitative improvement. Development in pedagogical awareness and reflective meaning-making was present but less stable, indicating the need for longer and more explicit scaffolding. The contribution of the study is therefore not a broad claim that digital collaborative learning automatically improves all outcomes. Its more precise contribution is to show how embedded assessment can function as narrative evidence of reasoning development in collaborative physics learning.

## CONCLUSION

This study shows that embedded assessment, when placed within a collaborative digital learning trajectory, can make students' reasoning in Simple Harmonic Motion more visible. The digital worksheets preserved students' initial ideas, their negotiated explanations during collaborative work, and their reflective consolidation at the end of instruction. Across the learning episodes, students' explanations moved from intuitive and descriptive accounts of repeated motion toward more relational reasoning involving displacement, restoring force, velocity, acceleration, and equilibrium. The quantitative evidence supported this pattern by showing improved conceptual understanding after students' initial ability was statistically controlled. However, the development of pedagogical awareness and reflective meaning-making remained less stable, suggesting that these dimensions require longer and more explicit scaffolding.

The main theoretical contribution of this study is the positioning of embedded assessment as narrative evidence of reasoning development, not merely as a mechanism for measuring final achievement. Practically, the findings suggest that digital worksheets are pedagogically valuable only when their prompts require students to predict, explain, justify, revise, and reflect within a coherent learning trajectory. For physics teacher education, this design offers a way to trace how pre-service teachers construct, question, and consolidate explanations of abstract physics concepts. Future studies should examine this approach across other physics topics, longer instructional cycles, and richer multimodal data, such as recorded group discussions, screen-based interaction records, and more detailed analyses of collaborative discourse.

## DECLARATION OF AI AND AI ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this manuscript, the author used Google NotebookLM and ChatGPT to assist in organizing literature notes, refining the manuscript structure, improving language clarity, and reviewing the coherence of arguments. These tools were not used to generate research data, conduct statistical analysis, or replace the author's academic judgment. All ideas, interpretations, data analysis, and scholarly claims were reviewed, verified, and revised by the author. The author takes full responsibility for the accuracy, integrity, and final content of the manuscript.

## ACKNOWLEDGEMENT

The author sincerely acknowledges the Ministry of Religious Affairs of the Republic of Indonesia and the Indonesia Endowment Fund for Education (LPDP) for providing financial support through the Beasiswa Indonesia Bangkit (BIB) scholarship program.

## AUTHORS' CONTRIBUTION

Author 1: Conceptualization; Methodology; Investigation; Data curation; Formal analysis
Author 2: Conceptualization; Supervision; Validation; Writing – review and editing.
Author 3: Methodology; Supervision; Validation; Writing – review and editing.
Author 4: Supervision; Validation; Writing – review and editing.
Author 5: Resources; Investigation support; Data organization

## DECLARATION OF COMPETING INTEREST

The authors declare that there are no financial, professional, or personal relationships that could have influenced the design, analysis, interpretation, or reporting of this manuscript.

## REFERENCES

Ainsworth, S. (2008). The Educational Value of Multiple-representations when Learning Complex Scientific Concepts. Dalam J. K. Gilbert, M. Reiner, & M. Nakhleh (Ed.), *Visualization: Theory and Practice in Science Education* (hlm. 191–208). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5267-5_9

Barkley, E. F., Major, C. H., & Cross, K. P. (2014). Collaborative learning techniques: A handbook for college faculty (2nd ed.). Jossey-Bass.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Bruner, J. S. (1990). *Acts of meaning*. Harvard University Press.

Clandinin, D. J., & Connelly, F. M. (2000). *Narrative inquiry: Experience and story in qualitative research* (1. ed., 1. PB print). Jossey-Bass.

Clark, A. K., & Karvonen, M. (2021). Instructionally Embedded Assessment: Theory of Action for an Innovative System. *Frontiers in Education*, *6*, 724938. https://doi.org/10.3389/feduc.2021.724938

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

Creswell, J. W., & Creswell, J. D. (2023). Research design: Qualitative, quantitative, and mixed methods approaches (6th ed.). SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.

Davidson, N. (2014). *Boundary Crossings: Cooperative Learning, Collaborative Learning, and Problem-Based Learning*.

Desimone, L. M., & Garet, M. S. (2015). Best Practices in Teachers' Professional Development in the United States. *Psychology, Society, & Education*, *7*(3), 252. https://doi.org/10.25115/psye.v7i3.515

DeVellis, R. F. (2017). Scale development: Theory and applications (4th ed.). SAGE Publications.

Dillenbourg, P. (1999). *What do you mean by "collaborative learning"?*

Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physical Review Special Topics - Physics Education Research*, *10*(2), 020119. https://doi.org/10.1103/PhysRevSTPER.10.020119

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman & Hall.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538. https://doi.org/10.1037/a0015808

Ferreira, A., & Lemmer, M. (2021). A fine-grained teaching activity to establish conceptual coherence in introductory physics concepts and representations. *Journal of Physics: Conference Series*, *1929*(1), 012003. https://doi.org/10.1088/1742-6596/1929/1/012003

Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (6th edition). Sage.

Hoehn, J. R., Gifford, J. D., & Finkelstein, N. D. (2020). *Epistemic stances toward group work in learning physics: Interactions between epistemology and social dynamics in a collaborative problem solving context* (arXiv:2005.02425). arXiv. https://doi.org/10.48550/arXiv.2005.02425

Holme, T. A., Luxford, C. J., & Brandriet, A. (2015). Defining Conceptual Understanding in General Chemistry. *Journal of Chemical Education*, *92*(9), 1477–1483. https://doi.org/10.1021/acs.jchemed.5b00218

Jacobsson, E., Wallenlind, S., & Pendrill, A.-M. (2024). Educational research in physics teacher education: The role of mathematics in physics teaching. *Physics Education*, *59*(6), 065021. https://doi.org/10.1088/1361-6552/ad7d57

Kim, B., & Reeves, T. C. (2007). Reframing research on learning with technology: In search of the meaning of cognitive tools. *Instructional Science*, *35*(3), 207–256. https://doi.org/10.1007/s11251-006-9005-2

Koerfer, E., & Gregorcic, B. (2024). Exploring student reasoning in statistical mechanics: Identifying challenges in problem-solving groups. *Physical Review Physics Education Research*, *20*(1), 010105. https://doi.org/10.1103/PhysRevPhysEducRes.20.010105

Laurillard, D. (2012). *Teaching as a design science: Building pedagogical patterns for learning and technology*. Routledge.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382–385. https://doi.org/10.1097/00006199-198611000-00017

McMillan, J. H. (2024). *Classroom assessment: Principles and practice that enhance student learning and motivation* (Eighth edition). Pearson Education, Inc.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (Edition 3). Sage.

Munfaridah, N., Avraamidou, L., & Goedhart, M. (2021). The Use of Multiple Representations in Undergraduate Physics Education: What Do we Know and Where Do we Go from Here? *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(1), em1934. https://doi.org/10.29333/ejmste/9577

Nugraha, D. A., Cari, C., Suparmi, A., & Sunarno, W. (2020). Investigation of Undergraduate Students Conceptual Understanding about Simple Harmonic Motion on Mass-Spring System. *Periódico Tchê Química*, *17*(35), 55–64. https://doi.org/10.52571/PTQ.v17.n35.2020.05_NUGRAHA_pgs_55_64.pdf

Odden, T. O. (2021). What Does It Mean to "Make Sense" of Physics? *The Physics Teacher*, *59*(8), 596–598. https://doi.org/10.1119/5.0024095

Odden, T. O. B., & Russ, R. S. (2019). Defining sensemaking: Bringing clarity to a fragmented theoretical construct. *Science Education*, *103*(1), 187–205. https://doi.org/10.1002/sce.21452

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health*, *29*(5), 489–497. https://doi.org/10.1002/nur.20147

Rochman, C., Mahen, E. C. S., & Nasrudin, D. (2018). Authentic Assessment Based on Teaching and Learning Trajectory with Student Activity Sheet (SAS) On Basic Physics Courses. *WaPFi (Wahana Pendidikan Fisika)*, *3*(1), 1. https://doi.org/10.17509/wapfi.v3i1.10373

Rochman, C., Nasrudin, D., & Kariadinata, R. (2017). Authentic Assessment Based on Teaching and Learning Trajectory: *2nd International Conference on Sociology Education*, 174–177. https://doi.org/10.5220/0007094901740177

Saldana, J. (2021). *The Coding Manual for Qualitative Researchers*.

Schiering, D., Sorge, S., Keller, M. M., & Neumann, K. (2023). A proficiency model for PRE-SERVICE physics teachers' pedagogical content knowledge ( PCK )—What constitutes HIGH-LEVEL PCK ? *Journal of Research in Science Teaching*, *60*(1), 136–163. https://doi.org/10.1002/tea.21793

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), 4. https://doi.org/10.2307/1175860

Somroob, S., & Wattanakasiwich, P. (2017). Investigating student understanding of simple harmonic motion. *Journal of Physics: Conference Series*, *901*, 012123. https://doi.org/10.1088/1742-6596/901/1/012123

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (Seventh edition). Pearson.

Ucan, S., & Webb, M. (2015). Social regulation of learning during collaborative inquiry learning in science: How does it emerge and what are its functions? International Journal of Science Education, 37(15), 2503–2532. https://doi.org/10.1080/09500693.2015.1083634

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, *43*(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396

W. Johnson, D., & T. Johnson, R. (2019). Cooperative Learning: The Foundation for Active Learning. Dalam S. Manuel Brito (Ed.), *Active Learning—Beyond the Future*. IntechOpen. https://doi.org/10.5772/intechopen.81086

Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), The future of assessment: Shaping teaching and learning (pp. 53–82). Routledge. https://doi.org/10.4324/9781315086545-3

Yusoff, M. S. B. (2019). ABC of Content Validation and Content Validity Index Calculation. *Education in Medicine Journal*, *11*(2), 49–54. https://doi.org/10.21315/eimj2019.11.2.6