JOURNAL OF EDUCATIONAL INNOVATION
IN SCIENCE AND TECHNOLOGY
Volume 1, Issue 1, June 2026, page: 25-30

DEVELOPMENT OF A DIGITAL TEST INSTRUMENT FOR THE INTEGRATED SCIENCE COURSE AT UNIVERSITAS SERAMBI MEKKAH

Mhd. Rufil Illiyin[1]*, Mahyana Mahyana[2], Muhammad Azzarkasyi[3]
[1,2,3]Physics Education, FKIP, Universitas Serambi Mekkah, Aceh, Indonesia
Email: [1]rufililliyin7@gmail.com; [2] Mahyana@serambimekkah.ac.id; [3]muhammadazzarkasyi@serambimekkah.ac.id

ABSTRACT

Digital assessment is increasingly used in higher education, but a platform is educationally useful only when its content, interface, and implementation are aligned with the course context. This study developed a Socrative-supported digital test instrument for the Integrated Science course at Universitas Serambi Mekkah and examined its initial feasibility. A Research and Development design comprised five stages: needs analysis, product design, expert validation, revision, and limited implementation. Forty students who had taken Integrated Science were selected purposively: 5 from Physics Education, 13 from Biology Education, and 22 from Chemistry Education. Data were collected using a needs-observation sheet, two 12-item expert-validation sheets, and a 15-item student-response questionnaire. All expert-validation items used a consistent five-point scale. The content expert awarded 57 of 60 points (95.00%), and the media expert awarded 53 of 60 points (88.33%); both results met the study's highly valid criterion. Student responses were positive for 16 students (40.0%), moderately positive for 18 (45.0%), and negative for 6 (15.0%). The product therefore showed strong initial content and media feasibility and was acceptable to most participants in limited use. These findings do not establish empirical item validity, test reliability, comparative effectiveness, or learning gains; those outcomes require broader psychometric and experimental evaluation.

Keywords: digital assessment; Integrated Science; Research and Development; Socrative; test instrument

1. INTRODUCTION

Assessment is not merely the final recording of a score; it is a process for obtaining evidence about what students know and can do in relation to intended learning outcomes. In an Integrated Science course, this requirement is especially important because the assessment must connect concepts from physics, biology, and chemistry rather than treat them as unrelated domains. Sound assessment design therefore requires alignment among the targeted knowledge, the tasks used to elicit evidence, and the interpretation of student responses (Pellegrino et al., 2001).

The expansion of digital learning has made online assessment a persistent trend in higher education. Reviews show that online assessment can support feedback, interaction, and the management of assessment evidence, but its educational value depends on clear criteria, instructional support, and purposeful design (Gikandi et al., 2011; Heil & Ifenthaler, 2023). The central issue is consequently not whether a quiz can be moved online, but whether the digital instrument preserves content relevance, understandable tasks, usable interaction, and defensible interpretation.

Research on e-assessment has also shown that digitization changes more than the medium used to deliver questions. It affects how tasks are presented, how responses are recorded, and how feedback can be returned to students (Stödberg, 2012). For a digital test to function as part of learning, its feedback must be timely, specific, understandable, and connected to the learner's response rather than limited to an unexplained score (Shute, 2008). These principles are particularly

* Corresponding author: rufililliyin7@gmail.com

Citation: Illiyin, M.R., Mahyana, M., & Azzarkasyi, M. (2026). Development of a Digital Test Instrument for the Integrated Science Course at Universitas Serambi Mekkah. *Journal of Educational Innovation in Science and Technology*, 1(1), 25–30.

relevant to Integrated Science because students must recognize relationships among concepts while also understanding why a selected response is appropriate.

Socrative is a web-based student-response platform through which instructors can administer quizzes, collect responses, and display results. Studies in physics and chemistry-related higher education have reported benefits associated with timely feedback, homework administration, and interactive formative assessment (Balta et al., 2018; Balta & Tzafilkou, 2019; Roman et al., 2021). A meta-analysis of audience-response systems nevertheless found that technological effects depend on instructional and research conditions, cautioning against attributing learning improvement to the platform alone (Hunsu et al., 2016). These findings support development and contextual evaluation before claims of effectiveness are made.

Platform acceptance is another condition that can influence implementation. In research on mobile-based assessment, perceived usefulness, ease of use, facilitating conditions, interface quality, and feedback were associated with students' intention to use the assessment system (Nikou & Economides, 2017). Socrative-specific studies likewise indicate that students' perceptions and classroom engagement can differ according to the response system and the way it is integrated into instruction (Abdulla, 2018; Guarascio et al., 2017). Comparative evaluation of digital classroom tools also shows that their perceived contributions to active learning, repetition, and feedback are not identical (Tautz et al., 2021). Consequently, expert validation of content and media should be considered together with students' responses during limited implementation.

Local evidence reinforces this need for contextual design. At Universitas Serambi Mekkah, Rizal et al. (2019) examined student perceptions of e-learning, making user acceptance a relevant consideration for technology-supported learning at the same institution. In an Aceh science-education context, Rizal et al. (2023) found that science teachers could develop online learning-outcome instruments while also reporting practical obstacles, including unfamiliarity with initial implementation and the additional preparation time required. The first study concerns e-learning perceptions and the second concerns school science teachers; neither directly validates a digital test for the university Integrated Science course. Their relevance lies in identifying local implementation conditions that a course-specific instrument should address.

Previous studies have mainly examined platform use, feedback, engagement, attitudes, or learning outcomes. Less attention has been given to the initial development of a course-specific digital test that integrates science content and separately documents content feasibility, media feasibility, and student responses in this institutional setting. The present study addresses that gap without assuming that favorable expert judgment or user response is equivalent to psychometric quality or causal effectiveness.

This study therefore asked: (1) How can a digital test instrument be developed for the Integrated Science course at Universitas Serambi Mekkah? (2) What levels of content and media feasibility are indicated by expert validation? and (3) How do students respond during limited implementation? The study argues that initial feasibility depends on the convergence of course alignment, usable media design, and user acceptance. Accordingly, the contribution is a context-specific product and bounded evidence for further testing, rather than a claim that digital delivery alone improves learning.

2. METHODS

This study used a Research and Development (R&D) design because its purpose was to produce a usable educational product and evaluate its initial feasibility. The procedure was adapted from educational product-development logic (Gall et al., 2003) and delimited to five stages: needs analysis, product design, expert validation, revision, and limited implementation. This delimitation was appropriate because the study focused on initial product feasibility, not large-scale dissemination or comparative effectiveness.

2.1 Setting and Participants

The study was conducted at the Faculty of Teacher Training and Education, Universitas Serambi Mekkah, from June 28 to August 28, 2025. The limited implementation involved 40 students selected purposively because they had taken the Integrated Science course and could evaluate the course-specific instrument. The sample comprised 5 Physics Education students (12.5%), 13 Biology

Education students (32.5%), and 22 Chemistry Education students (55.0%). The same total and subgroup counts were used throughout the analysis.

2.2 Development Procedure

Stage	Main activity	Output
1	Observation of assessment practice and identification of course-level needs	Product requirements
2	Preparation of the blueprint, questions, instructions, and Socrative configuration	Initial prototype
3	Separate review by one content expert and one media expert	Scores and revision notes
4	Revision of content, wording, presentation, navigation, and settings	Revised product
5	Limited administration and collection of student responses	User-response data

The needs analysis established the functional and content requirements. The product was then designed by aligning Integrated Science questions with the course scope and configuring their administration in Socrative. One content expert evaluated subject-matter alignment and item construction, while one media expert evaluated interface and operation. Their comments informed revisions before the product was administered to the 40 students. The sequence kept the design consistent with the reported results: product evidence preceded expert evidence, which preceded user-response evidence.

2.3 Instruments and Scoring

Data were collected using a needs-observation sheet, a content-validation sheet, a media-validation sheet, and a student-response questionnaire. Each expert sheet contained 12 statements. The content sheet addressed relevance to Integrated Science, conceptual accuracy, question construction, difficulty appropriateness, and meaningfulness. The media sheet addressed visual presentation, navigation, interactivity, access, and platform stability. The student questionnaire contained 15 statements on presentation, content, and perceived usefulness.

Score	Verbal anchor
5	Very good
4	Good
3	Adequate
2	Poor
1	Very poor

The same 1–5 scale was used for every expert-validation item. Expert percentages were calculated as P = (obtained score / maximum score) × 100. With 12 items, the maximum score for each expert sheet was 60. Percentages were interpreted using the study's operational intervals: highly valid (81–100%), valid (61–80%), moderately valid (41–60%), less valid (21–40%), and invalid (0–20%). Student responses were summarized using the three categories recorded in the response recapitulation—positive, moderately positive, and negative—and are reported only as frequencies and percentages. No inferential test or claim of learning improvement was made.

3. RESULTS

3.1 Needs Analysis and Product Development

The needs analysis identified the absence of a course-specific digital workflow that combined Integrated Science questions, device-based access, and automatic score recapitulation. These requirements were translated into a product containing user instructions, multiple-choice questions, response options, and a final score display. The initial prototype was reviewed before limited implementation. Revisions addressed the clarity of question wording and instructions, the presentation of content, navigation, and platform settings. The final prototype retained the Integrated Science focus while using Socrative as the delivery and scoring environment.

3.2 Expert Validation

Validator	Obtained score	Maximum score	Percentage	Category
Content expert	57	60	95.00%	Highly valid
Media expert	53	60	88.33%	Highly valid

The content expert awarded 57 of 60 points, corresponding to 95.00%. The media expert awarded 53 of 60 points, corresponding to 88.33%. Both percentages fell within the highly valid operational interval. The results provide evidence of content and media feasibility based on expert judgment. They do not establish empirical item validity or test-score reliability because those properties require student item-response data and psychometric analysis.

3.3 Student Responses

Study program	n	Positive	Moderately positive	Negative	Total
Physics Education	5	3	1	1	5
Biology Education	13	7	6	0	13
Chemistry Education	22	6	11	5	22
Total	40	16	18	6	40
Percentage	100%	40.0%	45.0%	15.0%	100%

Sixteen students (40.0%) gave positive responses, 18 (45.0%) gave moderately positive responses, and 6 (15.0%) gave negative responses. Thus, 34 of 40 students (85.0%) were in the positive or moderately positive categories. The finding describes initial acceptance of the product during limited implementation; it is not a measure of academic achievement.

4. DISCUSSION

The study produced a digital test through a coherent sequence from needs identification to limited use. The product addressed the practical need for course-specific questions, student-device access, and automated scoring. More importantly, it treated Socrative as the delivery environment rather than as a substitute for assessment design. This distinction follows the assessment principle that tasks and interpretations must remain aligned with the knowledge domain being assessed (Pellegrino et al., 2001).

The content score of 95.00% and media score of 88.33% indicate that the prototype met the predetermined feasibility criterion from two complementary perspectives. The higher content score suggests particularly strong expert agreement with the course alignment and construction of the questions, while the media score indicates a favorable but comparatively lower appraisal of presentation and operation. Because each domain was judged by one expert, these percentages should be read as initial expert evidence, not consensus estimates or psychometric validation.

The product's feedback and response-management functions are consistent with the wider online-assessment literature. Systematic and narrative reviews identify feedback, clear criteria, and instructional support as recurring conditions for effective online assessment (Gikandi et al., 2011; Heil & Ifenthaler, 2023). Socrative studies in physics and chemistry-related courses similarly show that the platform can support instant feedback, homework, and interactive quizzes when it is embedded in a deliberate instructional design (Balta et al., 2018; Balta & Tzafilkou, 2019; Roman et al., 2021). The present result extends this literature only at the product-feasibility level; it does not replicate their outcome-based designs.

The meaning of automatic scoring also requires qualification. Rapid score display can reduce administrative work, but a score becomes educationally informative only when students can connect it with the quality of their responses. Stödberg (2012) identified feedback and assessment design as central concerns in e-assessment research, while Shute (2008) showed that formative feedback is most useful when it is timely, specific, and directed toward improvement. The developed instrument provides the technical basis for prompt response information, but the present data do not establish the quality of feedback explanations or their effect on subsequent learning. A later version should therefore connect score reporting with item-level explanations and opportunities for revision.

The 85.0% combined positive and moderately positive response indicates broad initial acceptance, whereas the 15.0% negative response shows that acceptance was not uniform. This mixed pattern is meaningful in the local context. Rizal et al. (2019) positioned student perception as an implementation issue for e-learning at Universitas Serambi Mekkah, while Rizal et al. (2023) documented practical difficulties and preparation demands in the development of online science assessment instruments. Taken together, these studies help explain why content quality, usability, and implementation support must be considered simultaneously. They do not, however, provide direct evidence about the participants or product in the present study.

The response distribution can also be interpreted through technology-acceptance conditions rather than as a simple approval rating. Nikou and Economides (2017) found that usefulness, facilitating conditions, interface, feedback, and related factors contributed to intention to use mobile-based assessment. Guarascio et al. (2017) similarly compared student perceptions of Socrative with a traditional response system, and Abdulla (2018) examined its role in supporting learning during university physiology lectures. These studies suggest plausible reasons why most students accepted the instrument while a minority responded negatively, including differences in perceived usefulness, familiarity, access, and interaction. Because the present questionnaire results were aggregated, these explanations remain literature-informed interpretations and should be tested with item-level response data.

The expert and student evidence should therefore be read as complementary. Expert validation addressed whether the content and media met predetermined development criteria, whereas student responses indicated whether the revised product was acceptable in actual limited use. Neither source of evidence can replace the other. A technically usable platform may contain poorly aligned questions, while strong questions may still be difficult to access or navigate. This distinction is consistent with research showing that digital response systems support engagement and learning only when their affordances are aligned with pedagogy and classroom practice (Guarascio et al., 2017; Hunsu et al., 2016).

The study contributes a bounded development model for Integrated Science: define the course need, align questions with the integrated domain, validate content and media separately, revise the product, and then examine user response. This sequence is useful because it prevents platform convenience from being mistaken for assessment quality. It also reveals the next evidence required: item-level response analysis, reliability estimation, broader validation, and comparison with alternative assessment procedures.

Several limitations constrain the claims. The implementation involved a purposive sample from one institution. Content and media were each reviewed by one expert, so interrater agreement could not be estimated. The available analysis did not include item difficulty, discrimination, empirical validity, internal-consistency reliability, completion time, or learning outcomes. The student-response results were aggregated into three categories, limiting item-level interpretation. Future research should involve multiple validators, retain item-level validation and response data, test the instrument with a larger and more diverse sample, and conduct psychometric and comparative analyses.

5. CONCLUSION

A Socrative-supported digital test instrument for the Integrated Science course at Universitas Serambi Mekkah was developed through needs analysis, product design, expert validation, revision, and limited implementation. The content-validation result was 95.00% and the media-validation result was 88.33%, both within the study's highly valid interval. Among 40 purposively selected students, 40.0% responded positively, 45.0% moderately positively, and 15.0% negatively. The converging evidence supports the product's initial feasibility for limited use after revision. Its contribution is a context-specific assessment product and a transparent basis for further testing. Claims about empirical item validity, reliability, comparative effectiveness, and learning gains remain outside the evidence generated by this study.

REFERENCES

Abdulla, M. H. (2018). The use of an online student response system to support learning of physiology during lectures to medical students. *Education and Information Technologies*, 23(6), 2931–2946. https://doi.org/10.1007/s10639-018-9752-0

Balta, N., Perera-Rodríguez, V.-H., & Hervás-Gómez, C. (2018). Using Socrative as an online homework platform to increase students' exam scores. *Education and Information Technologies*, 23, 837–850. https://doi.org/10.1007/s10639-017-9638-6

Balta, N., & Tzafilkou, K. (2019). Using Socrative software for instant formative feedback in physics courses. *Education and Information Technologies*, 24(1), 307–323. https://doi.org/10.1007/s10639-018-9773-8

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Allyn & Bacon.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57(4), 2333–2351. https://doi.org/10.1016/j.compedu.2011.06.004

Guarascio, A. J., Nemecek, B. D., & Zimmerman, D. E. (2017). Evaluation of students' perceptions of the Socrative application versus a traditional student response system and its impact on classroom engagement. *Currents in Pharmacy Teaching and Learning*, 9(5), 808–812. https://doi.org/10.1016/j.cptl.2017.05.011

Heil, J., & Ifenthaler, D. (2023). Online assessment in higher education: A systematic review. *Online Learning*, 27(1), 187–218. https://doi.org/10.24059/olj.v27i1.3398

Hunsu, N. J., Adesope, O., & Bayly, D. J. (2016). A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. *Computers & Education*, 94, 102–119. https://doi.org/10.1016/j.compedu.2015.11.013

Nikou, S. A., & Economides, A. A. (2017). Mobile-based assessment: Investigating the factors that influence behavioral intention to use. *Computers & Education*, 109, 56–73. https://doi.org/10.1016/j.compedu.2017.02.005

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press. https://doi.org/10.17226/10019

Rizal, S., Kasmawati, Walidain, B., Wahyuni, A., & Musdar. (2019). Student perception toward e-learning in Primary School Teacher Education Department students of Serambi Mekkah University. *Journal of Physics: Conference Series*, 1232(1), Article 012053. https://doi.org/10.1088/1742-6596/1232/1/012053

Rizal, S., Azzarkasyi, M., Firmansyah, J., & Ibrahim. (2023). Perceptions of science teachers in developing student learning outcomes instruments online. *Jurnal Penelitian Pendidikan IPA*, 9(9), 6957–6961. https://doi.org/10.29303/jppipa.v9i9.4161

Roman, C., Delgado, M. A., & García-Morales, M. (2021). Socrative, a powerful digital tool for enriching the teaching–learning process and promoting interactive learning in chemistry and chemical engineering studies. *Computer Applications in Engineering Education*, 29(6), 1542–1553. https://doi.org/10.1002/cae.22408

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. https://doi.org/10.3102/0034654307313795

Stödberg, U. (2012). A research review of e-assessment. *Assessment & Evaluation in Higher Education*, 37(5), 591–604. https://doi.org/10.1080/02602938.2011.557496

Tautz, D., Sprenger, D. A., & Schwaninger, A. (2021). Evaluation of four digital tools and their perceived impact on active learning, repetition and feedback in a large university class. *Computers & Education*, 175, Article 104338. https://doi.org/10.1016/j.compedu.2021.104338