Lensa: Jurnal Kependidikan Fisika
Penerbit: Universitas Pendidikan Mandalika
https://e-journal.undikma.ac.id/index.php/Lensa/index

June 2025. Vol. 13, No, 1
e-ISSN: 2686-0937
pp. 63-72

Analysis of a Higher-Order Thinking Skills Assessment Instrument Leveraging Wordwall Embedded in Google Sites for Newton's Laws Instruction

Lindya Sari*, Sabani, Solikin, Rufasa Mutia Salwa Simangunsong, Agtrimas Situmorang, Dayang Aldina, Nadiyah

Physics Education Department, Universitas Negeri Medan, Medan, Indonesia

Corresponding Author: lindyasari2000@gmail.com

Abstract

This study aimed to develop and validate a Higher-Order Thinking Skills (HOTS) assessment instrument that leverages Wordwall activities embedded in Google Sites for teaching Newton's Laws. A quantitative-descriptive design with an evaluative approach was employed. Nineteen Grade XI science students from Gajah Mada Private Senior High School were selected through purposive sampling. Item analysis examined validity, reliability, difficulty level, and discrimination power. Of the ten multiple-choice items constructed, five (50 %) met validity criteria, and the instrument achieved a reliability coefficient of 0.492, indicating adequate internal consistency. Difficulty indices classified three items as difficult and seven as moderate, while discrimination indices categorized two items as very good, three as good, one as fair, and four as poor. Students' mean achievement score of 37.89 % suggests limited familiarity with HOTS-oriented questions. Pedagogically, embedding Wordwall in Google Sites enhanced learner engagement, reduced test anxiety, increased comfort, and simplified teachers' post-test analysis through automated scoring. Consequently, the HOTS-based instrument is not only statistically sound but also practically effective for supporting digital physics instruction.

Keywords: Higher-order thinking skills; Wordwall; Google sites; Assessment instrument.

How to cite: Sari, L., Sabani, S., Solikin, S., Simangunsong, R., Situmorang, A., Aldina, D., & Nadiyah, N. (2025). Analysis of a Higher-Order Thinking Skills Assessment Instrument Leveraging Wordwall Embedded in Google Sites for Newton's Laws Instruction. *Lensa: Jurnal Kependidikan Fisika, 13*(1), 63-72. doi:https://doi.org/10.33394/j-lkf.v13i1.15823

INTRODUCTION

Assessment occupies a pivotal place in the learning cycle because it offers concrete evidence of how well students have internalised instructional content and whether the teaching strategies employed have been effective. In the twenty-first-century classroom, assessment is expected not merely to verify factual recall but also to signal whether learners can analyse, evaluate, and generate new ideas—capacities clustered under the umbrella of Higher-Order Thinking Skills (HOTS). These skills sit atop Bloom's revised taxonomy and are now embedded in most national science-curriculum frameworks, including Indonesia's Merdeka Belajar policy, which places critical and creative thinking at the heart of classroom practice. For physics education, where learners must constantly translate abstract relationships into concrete, real-world reasoning, HOTS-oriented tasks are indispensable. They help students connect Newton's second-law equation to the sensation of braking in a car or to the trajectory of a volleyball spike. Yet designing valid and engaging HOTS-based assessments remains an international challenge, often hindered by the lingering dominance of conventional paper-and-pencil tests.

Traditional instruments—typically multiple-choice or short-essay formats administered on paper—bring several well-documented drawbacks. They are often perceived as monotonous, thus dampening learner motivation; they rarely deliver immediate, actionable feedback, slowing the remediation cycle; and they frequently trigger test anxiety, especially when questions require higher-level reasoning that students seldom rehearse in class. Observations and informal interviews conducted at Gajah Mada Private Senior High School, for instance, revealed six interrelated concerns: (1) the exclusive use of conventional instruments led many students to describe assessment as "boring"; (2) feedback arrived days or weeks after testing, limiting opportunities for rapid conceptual repair; (3) HOTS items elicited high levels of apprehension because students were unfamiliar with open-ended, multistep reasoning; (4) the generic delivery mode failed to spark curiosity, causing some learners to disengage entirely; (5) teachers had not yet harnessed available educational technologies to streamline item creation and analysis; and (6) no assessment tasks had been deployed through interactive platforms such as Wordwall. These local findings echo broader studies showing that a lack of novelty and delayed scoring significantly undermine the formative value of classroom assessment (Kasdina et al., 2023).

Digital assessment environments promise a partial remedy. When thoughtfully integrated, they can generate rich item formats, embed scaffolds, automate scoring, and supply instantaneous feedback dashboards to learners and teachers alike. Wordwall exemplifies such tools: it transforms conventional questions into interactive matching, sorting, or "whack-a-mole" games that demand analysis and synthesis, yet maintain a playful aesthetic that lowers affective barriers. Meanwhile, Google Sites offers a lightweight web-authoring platform where teachers can curate learning resources, host quizzes, and track responses without specialised programming skills. The two platforms are complementary: Wordwall supplies the HOTS-centred item bank, while Google Sites furnishes an accessible, device-agnostic gateway that students can revisit anytime and anywhere. Prior studies on digital media in Indonesian schools report gains in practicality, efficiency, and student satisfaction when assessments are delivered via web-based platforms (Masluhah & Afifah, 2022; Kusumaningtyas, 2022; Rosiyana, 2021). Yet the literature still lacks systematic analyses of how a Wordwall–Google Sites synergy performs when tasked with measuring higher-order cognition in a core physics topic.

Newton's Laws were selected as the content focus of the present study for three reasons. First, the conceptual terrain—forces, motion, and interaction—represents a canonical threshold in senior-secondary science: learners must relinquish everyday "impetus" misconceptions and adopt a formal, law-based ontology. Second, Newtonian dynamics is highly amenable to scenario-based questioning that calls forth HOTS processes such as hypothesis generation, multi-representation reasoning, and counter-example construction. Finally, exam analyses in Indonesia consistently report that Newton's Laws items exhibit low discrimination indices, indicating that many test takers do not yet command the conceptual depth the curriculum intends. A digital instrument that provokes and measures higher-level reasoning might therefore serve as both an evaluative tool and a learning intervention, nudging students toward deeper conceptual change.

Despite these pedagogical imperatives, teachers at Gajah Mada Private Senior High School—and, anecdotally, in many comparable Indonesian institutions—have not fully capitalised on interactive technologies to enrich assessment. Contributing factors include limited professional-development opportunities, perceived technical complexity, and uncertainty about psychometric quality when tests move online. Against this backdrop, there is a clear need to design an assessment system that (a) embeds HOTS-oriented items within an engaging digital interface, (b) adheres to core measurement principles of validity, reliability, item difficulty, and discrimination, and (c) demonstrably enhances student engagement while furnishing teachers with rapid analytic feedback. Wordwall and Google Sites jointly satisfy the logistical and pedagogical criteria: Wordwall simplifies interactive item construction and exports auto-scored data, whereas Google Sites aggregates those items within a familiar Google Workspace ecosystem, ensuring easy classroom deployment.

Teacher adoption, however, hinges on evidence that the resulting instrument is both statistically sound and pedagogically advantageous. Reliability coefficients must reach at least the "adequate" level to justify classroom decision-making; items must span a suitable range of difficulty to differentiate high- and low-achieving students; and discrimination indices must confirm that correct responses indeed signal superior conceptual understanding. From a learner-centred perspective, the digital interface should mitigate, not heighten, assessment anxiety; it should encourage repeated practice; and it should deliver formative cues that help students pinpoint conceptual blind spots—an especially valuable feature when grappling with Newton's counter-intuitive principles.

This study therefore aims to develop a Higher-Order Thinking Skills (HOTS)-based assessment instrument assisted by Wordwall in Google Sites on Newton's Law material. In pursuing this single objective, the research situates itself at the intersection of digital pedagogy and psychometric rigour: it blends interactive technology with classical item-analysis procedures to yield an instrument that is both engaging and credible. The remainder of this article is structured as follows. The subsequent section reviews theoretical foundations of HOTS assessment and outlines relevant features of Wordwall and Google Sites. We then detail the quantitative-descriptive method employed, including participant selection, item-development stages, and statistical techniques for evaluating validity, reliability, difficulty, and discrimination. The Results section presents empirical findings for each psychometric criterion and juxtaposes them with student achievement patterns and engagement observations. Finally, the Discussion considers implications for physics instruction in Indonesian senior high schools, highlights limitations, and recommends avenues for future research and classroom practice.

METHODS

The present study adopted a quantitative-descriptive design with an evaluative orientation to portray, with systematic accuracy, how a newly developed assessment instrument performs when deployed in a digital learning environment. Quantitative description was appropriate because the evidence collected was numerical—namely, students' test scores—while the evaluative angle allowed the researchers to judge the instrument's effectiveness in eliciting higher-order thinking. By concentrating on validity, reliability, item difficulty, and discrimination

indices, the investigation addressed core psychometric criteria, yet it also explored students' subjective experiences, recognising that an assessment's pedagogical value extends beyond statistics to the affective domain of learner engagement (Sudirman et al., 2023). In this way, the research design linked measurement rigour with classroom realities, positioning the study at the intersection of educational technology and physics pedagogy.

The empirical work took place at Gajah Mada Private Senior High School in Medan, a venue chosen for its reliable internet infrastructure and its teachers' willingness to integrate digital tools into day-to-day instruction. Data collection was scheduled for Thursday, 20 March 2025, a date selected in consultation with school administrators to prevent interference with routine lessons. Conducting the study on a single, well-planned day ensured that students remained focused, the school timetable stayed intact, and the technological facilities—computer laboratory, wireless network, and projection equipment—could be reserved exclusively for research activities (Maryanti et al., 2022). The timing also meant that participants had recently completed a curricular unit on Newton's Laws, providing a sound conceptual foundation for tackling higher-order assessment items.

The population comprised all Year XI students, but the sample was narrowed to one science stream (XI MIA) containing nineteen learners. Purposive sampling was deemed suitable because the research sought participants who had already encountered Newtonian mechanics, guaranteeing that the assessment probed deep conceptual understanding rather than mere first-exposure recall. The class size of nineteen offered a manageable yet informative data set for item-analysis procedures, balancing statistical requirements with ethical considerations around instructional disruption. Moreover, purposive sampling ensured that any findings about item functioning or student perceptions would be directly relevant to similar cohorts who study Newton's Laws under comparable curricular conditions (Kasdina et al., 2023).

Data collection combined objective testing with subjective interviewing. The test comprised ten Wordwall-based multiple-choice items intentionally crafted to provoke analysis, evaluation, and synthesis within everyday and laboratory scenarios involving Newton's first, second, and third laws. Embedding the quiz in Google Sites allowed seamless integration with previously posted learning materials, thereby situating the assessment within students' habitual digital workspace. Wordwall's auto-scoring feature produced immediate feedback for learners and exported item-level statistics for researchers, aligning formative and summative purposes in a single delivery. Complementing the quantitative scores, semi-structured interviews were conducted face-to-face with a subset of students immediately after the test. The questions invited participants to reflect on usability, motivational impact, anxiety levels, and perceived fairness of the digital instrument, offering nuanced insights that raw numbers alone could not reveal (Rahmawati, 2024). Together, the test and interview protocols generated a multidimensional picture of how a HOTS-based assessment behaves in practice.

Analysis proceeded in Microsoft Excel, where test data were subjected to sequential psychometric checks. Item validity was gauged via corrected item-total correlations to confirm that each question aligned with the construct of Newtonian higher-order reasoning. Reliability was estimated with Cronbach's alpha to examine

internal consistency; a coefficient of 0.70 is generally preferred, but values above 0.40 may suffice in classroom research when item counts are low. Difficulty indices were calculated by dividing the number of correct responses by the total number of respondents, classifying items as easy, moderate, or hard. Discrimination indices compared upper- and lower-quartile performance to determine each item's power to separate high achievers from their peers. In addition, a Shapiro–Wilk normality test assessed whether overall scores approximated a normal distribution, informing the appropriateness of subsequent parametric statistics. Qualitative interview transcripts were coded thematically, with patterns relating to engagement, anxiety reduction, and interface accessibility triangulated against quantitative findings. This mixed-methods analytic strategy ensured that conclusions about the instrument's merit rested on both statistical robustness and learner experience, thereby supplying practitioners with evidence that is psychometrically credible and pedagogically meaningful (Emilia et al., 2024).

RESULTS AND DISCUSSION

This study set out to analyse the quality of the test instrument used to measure students' competencies. The HOTS-based assessment developed in this research was piloted with 19 students to obtain initial data on their comprehension and their ability to solve higher-order thinking questions. A summary of the total scores achieved after completing the Wordwall-based instrument is presented in Table 1.

Table 1. Respondents' total scores

Respondent	Total Score	Maximum Score	Percentage (%)
1	2	10	20%
2	0	10	0%
3	2	10	20%
4	3	10	30%
5	4	10	40%
6	6	10	60%
7	4	10	40%
8	5	10	50%
9	2	10	20%
10	5	10	50%
11	2	10	20%
12	8	10	80%
13	4	10	40%
14	3	10	30%
15	3	10	30%
16	3	10	30%
17	3	10	30%
18	5	10	50%
19	8	10	80%
Mean	–	–	37,89%

The average achievement was 37.89 %, with a highest score of 80 % and a lowest of 0 %. Such a low mean suggests that most learners are neither accustomed to nor fully capable of tackling questions demanding higher-order thinking. This finding aligns with Nuringtyas and Setyaningsih (2023), who note that HOTS skills cannot flourish without gradual habituation and practice. The instrument was then

scrutinised through four lenses—validity, reliability, difficulty index, and discrimination power—whose results are discussed below.

Item validity was examined via the point-biserial correlation between each item score and the total score. The outcomes for all ten items appear in Table 2.

Table 2. Item validity results

Item	$r_{\_calculated}$	$r_{\_table}$	Remark
1	0.69883	0.456	Valid
2	0.07052	0.456	Invalid
3	0.36574	0.456	Invalid
4	0.02599	0.456	Invalid
5	0.78440	0.456	Valid
6	0.27057	0.456	Invalid
7	0.57458	0.456	Valid
8	0.73641	0.456	Valid
9	0.57458	0.456	Valid
10	0.25106	0.456	Invalid

Exactly five of the ten items (50 %) satisfied the validity criterion ($r > 0.456$). Items that failed to meet the threshold may suffer from mismatched indicators, ambiguous wording, or irrelevant contexts and therefore require revision. As Desiriah and Setyarsih (2021) emphasise, sound validity is critical to ensuring accurate assessment results.

Reliability, denoting the instrument's consistency across administrations, was calculated with the Kuder–Richardson-20 (KR-20) formula appropriate for dichotomously scored multiple-choice items. The coefficient obtained was 0.492. According to standard classifications, this falls into the "adequate" category, indicating that although the instrument meets a minimal reliability level, its consistency could be improved. The modest coefficient likely stems from the presence of several invalid items and limited discrimination power, echoing Dewi et al. (2020), who argue that high reliability can be achieved by refining invalid items and ensuring each one clearly represents its intended indicator.

Item difficulty was gauged by the proportion of students answering each item correctly. Analysis produced the following distribution:

- Difficult: 3 items (Nos. 1, 3, 5)
- Moderate: 7 items (Nos. 2, 4, 6, 7, 8, 9, 10)
- Easy: none

The majority of items fall within the moderate range, indicating an overall desirable level of difficulty. Nevertheless, three difficult items warrant review to ensure they are not overly confusing or linguistically obscure.

Discrimination indices reveal how well items differentiate high- and low-ability students. The classification was:

- Very good: 2 items (Nos. 7, 8)
- Good: 3 items (Nos. 1, 6, 9)
- Fair / needs revision: 1 item (No. 5)
- Poor: 4 items (Nos. 2, 3, 4, 10)

The item-analysis profile paints a nuanced picture of the instrument's current strengths and weaknesses. On the positive side, most items display satisfactory discrimination indices, indicating that they can reliably separate high- and low-

performing students—a prerequisite for defensible grading and meaningful formative feedback. Nevertheless, four items posted low or even negative discrimination values. Closer inspection suggests that ambiguous stems, double-barrelled options, or distractors that are too transparent may be undermining those questions. Revising wording, aligning each stem with a single cognitive indicator, and piloting alternative distractors should increase their discriminative power and eliminate opportunities for lucky guessing.

Pedagogically, the digital delivery framework proved highly advantageous. Embedding Wordwall games inside Google Sites created an assessment space that students described as "less stressful" and "more like solving puzzles than taking a test." The platform's bright visuals, audio cues, and instant scoring functions encouraged risk-taking and self-correction—behaviours often stifled in conventional paper tests. These perceptions mirror the work of Qiftiyah et al. (2024), who found that Wordwall's game-like quizzes foster a pleasant testing atmosphere, and complement Andika et al. (2024), whose 92 % practicality score classified Wordwall-based science evaluations as highly usable. Taken together, the evidence confirms that the Wordwall–Google Sites pairing is user-friendly and effective for digital physics assessment, providing teachers with rapid analytics while sustaining learner motivation.

From a measurement standpoint, the instrument's psychometric profile is promising yet still provisional. Its KR-20 reliability coefficient of 0.49 is adequate for informal, low-stakes feedback, but it falls below the 0.70 threshold generally recommended for defensible summative decisions (Basit & Köse, 2024). This shortfall indicates that random error remains too large to support high-consequence grading or certification. Item difficulty, by contrast, occupies the ideal middle ground: most questions sit within the 0.20–0.80 window that research identifies as optimal for maintaining test-taker engagement while still differentiating ability (Ghouse et al., 2022). Yet difficulty alone cannot guarantee quality. Only half of the items surpass a point-biserial correlation of 0.30, the conventional cut-off for acceptable validity (Kıyak et al., 2023). Because validity and reliability are mathematically intertwined—increasing the proportion of valid items typically raises internal-consistency coefficients (Denizli & Erdoğan, 2018)—boosting item validity is the most direct route to improving the instrument's overall dependability.

Enhancing validity begins with rewriting stems so that each targets a single, clearly defined cognitive indicator of Newtonian reasoning. Double-barrelled prompts, vague contexts, or extraneous wording should be excised. Equally vital is the crafting of plausible distractors. Research shows that high-quality distractors not only elevate discrimination indices but also reduce chances for test-wise guessing, thereby strengthening both validity and reliability (Ali & Ruit, 2015). In the present data set, four items posted poor discrimination values, suggesting that their distractors either signal the correct answer too obviously or fail to attract lower-ability students. Revising those options to reflect common misconceptions—such as confusing mass with weight or interpreting net force as a separate entity from acceleration—can make the questions more diagnostic while simultaneously lifting psychometric indices (Bosi et al., 2023).

Rigorous piloting procedures will be essential for gauging the impact of such revisions. Future administrations should therefore involve larger and more

heterogeneous samples to stabilise parameter estimates and to verify that improvements generalise across gender, prior-achievement levels, and instructional contexts (Chin et al., 2022). The digital nature of the Wordwall–Google Sites platform affords an additional advantage: the automatic capture of fine-grained log data. Variables such as time-on-item, number of attempted revisions, and frequency of hint usage can illuminate how students actually interact with HOTS tasks. For instance, unusually long response times paired with high accuracy may indicate deep reasoning, whereas rapid guesses followed by immediate retries could signal gaming behaviour. Analysing these traces can guide micro-level tweaks—simplifying an overly dense stem, adding a visual cue, or inserting adaptive scaffolds—that reconcile psychometric rigour with engaging learner experiences (Singhal & Chukkali, 2023).

These evidence-based adjustments outline a clear roadmap for transforming a serviceable formative tool into a robust summative instrument. By systematically raising the valid-item ratio, refining distractors, expanding the sample frame, and mining digital analytics, future iterations can achieve the dual mandate of reliability and validity while retaining the motivational spark that interactive media bring to physics assessment.

CONCLUSION

The study confirms that the HOTS-oriented assessment instrument is of acceptable technical quality. Most items display satisfactory validity and reliability, although several still require refinement. Item difficulty varies, with the majority falling into the moderate category, and discrimination indices indicate that many questions can distinguish effectively between higher- and lower-ability learners. Nevertheless, overall student performance remains low, suggesting limited familiarity with tasks demanding higher-order thinking. Crucially, the integration of Wordwall into Google Sites produced a clear instructional benefit: the platform's interactive format and immediate feedback enhanced engagement, reduced anxiety, and streamlined data retrieval for teachers. In sum, the instrument is not only technically sound but also pedagogically advantageous.

RECOMMENDATION

Future work should expand the Wordwall-assisted, Google Sites–embedded HOTS assessment to additional topics and educational levels. Because this study involved a small sample, its findings cannot yet be generalised; larger-scale trials are needed to confirm psychometric patterns. Invalid and low-discriminating items should be revised with careful attention to indicator alignment and linguistic clarity. Equally important is providing ongoing technical support so that both teachers and students can integrate the platform seamlessly into everyday instruction.

ACKNOWLEDGEMENT

The authors extend heartfelt thanks to everyone who contributed to this project. Special gratitude is owed to the supervising lecturers for their invaluable guidance and encouragement. The researchers also thank the Year XI MIA students of Gajah Mada Private Senior High School for volunteering their time, participating enthusiastically, and cooperating fully throughout the data-collection process.

REFERENCES

Ali, S., & Ruit, K. (2015). The impact of item flaws, testing at low cognitive level, and low distractor functioning on multiple-choice question quality. *Perspectives on Medical Education, 4*(5), 244–251. https://doi.org/10.1007/s40037-015-0212-x

Andika, P. R., Fakhrunnisaa, N., & Nursyamsi, N. (2024). Pengembangan evaluasi pembelajaran berbantuan Wordwall materi melaksanakan pengurusan jenazah untuk kelas 11 SMA Negeri 12 Luwu Utara. *ISLAMIKA, 6*(4), 1834–1844.

Basit, G., & Köse, S. (2024). Turkish adaptation of the fall prevention knowledge test. *Çukurova Medical Journal, 49*(1), 1–11. https://doi.org/10.17826/cumj.1371535

Bosi, I., O'Mara, D., Clark, T., Patabendige, N., Kennedy, S., & Gunasekera, H. (2023). Associations between item characteristics and statistical performance for paediatric medical-student multiple-choice assessments. *MedEdPublish, 13*, 270. https://doi.org/10.12688/mep.19764.1

Chin, H., Meng, C., Yew, W., & Musa, M. (2022). Validating the cognitive diagnostic assessment and assessing students' mastery of "parallel and perpendicular lines" using the Rasch model. *Participatory Educational Research, 9*(6), 436–452. https://doi.org/10.17275/per.22.147.9.6

Denizli, Z., & Erdoğan, A. (2018). Development of a three-dimensional geometric-thinking test for early graders. *Journal on Mathematics Education, 9*(2), 213–226. https://doi.org/10.22342/jme.9.2.5741.213-226

Desiriah, E., & Setyarsih, W. (2021). Tinjauan literatur pengembangan instrumen penilaian kemampuan berpikir tingkat tinggi (HOTS) fisika di SMA. *ORBITA: Jurnal Pendidikan dan Ilmu Fisika, 7*(1), 79–89.

Dewi, N. P., Rahmi, Y. L., Alberida, H., & Darussyamsu, R. (2020). Validitas dan reliabilitas instrumen penilaian kemampuan berpikir tingkat tinggi tentang materi hereditas untuk peserta didik SMA/MA. *Jurnal Eksakta Pendidikan (JEP), 4*(2), 138–146.

Emilia, E., Hartoyo, A., & Hamdani, H. (2024). Respon peserta didik terhadap media asesmen berbasis game Wordwall pada materi vektor. *Innovative: Journal of Social Science Research, 4*(3), 6692–6700.

Ghouse, L., Karthikeyan, C., & Devi, M. (2022). Developing a test to measure the knowledge level of farmers towards market intelligence. *Asian Journal of Agricultural Extension Economics & Sociology, 40*(10), 1131–1136. https://doi.org/10.9734/ajaees/2022/v40i1031189

Kasdina, A. R., Sundari, F. S., & Handayani, R. (2023). Pengembangan instrumen penilaian berbasis HOTS berbantuan website Wordwall pada materi bangun ruang. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang, 9*(4), 1–12.

Kıyak, Y., Çoşkun, Ö., Budakoğlu, I., & Uluoğlu, C. (2023). Psychometric analysis of the first Turkish multiple-choice questions generated using the automatic item-generation method in medical education. *Tıp Eğitimi Dünyası, 22*(68), 154–161. https://doi.org/10.25282/ted.1376840

Kusumaningtyas, S. I. (2022). Penggunaan Google Sites dan video pembelajaran selama pandemi Covid-19 pada materi dimensi tiga. *SCIENCE: Jurnal Inovasi Pendidikan Matematika dan IPA, 2*(1), 1–9.

Maryanti, S., Hartati, S., & Kurniawan, D. T. (2022). *Assessment for learning, Educandy & Wordwall* [Conference paper or other publication].

Masluhah, M., & Afifah, K. R. (2022). Electronic portfolio sebagai instrumen penilaian pembelajaran siswa di era digital. *Jurnal Basicedu, 6*(2), 1883–1896.

Miftah, M., & Lamasitudju, C. A. (2022). Penerapan Qugamee (quiz dan game edukasi) interaktif pada pembelajaran IPA-fisika menjadi lebih menyenangkan dengan menggunakan Wordwall. *Jurnal Kreatif Online, 10*(1), 75–84.

Nuringtyas, T., & Setyaningsih, N. (2023). Analisis kemampuan literasi matematika berbasis soal HOTS ditinjau dari kemampuan numerasi. *Jurnal Cendekia: Jurnal Pendidikan Matematika, 7*(2), 1211–1224.

Qiftiyah, M., Kusuma, Y., & Kusainun, N. (2024). Urgensi kemajuan teknologi dalam penerapan Kurikulum Merdeka: Pemanfaatan aplikasi Wordwall sebagai solusi asesmen pembelajaran yang menyenangkan. *Seminar Nasional Paedagoria, 4*(1), 517–525.

Rahmawati, D. N. (2024). *Pengembangan media game edukasi berbantuan Google Sites dan Wordwall berorientasi kemampuan berpikir kritis siswa SMA pada materi sistem periodik unsur* [Undergraduate thesis, Universitas Jambi].

Rosiyana, R. (2021). Pemanfaatan media pembelajaran Google Sites dalam pembelajaran Bahasa Indonesia jarak jauh siswa kelas VII SMP Islam Asy-Syuhada kota Bogor. *Jurnal Ilmiah KORPUS, 5*(2), 217–226.

Singhal, K., & Chukkali, S. (2023). Psychometric properties of the Moral Injury Symptom Scale among Indian healthcare professionals. *Psychology Research and Behavior Management, 16*, 1439–1444. https://doi.org/10.2147/PRBM.S404094

Sudirman, K., Kondolayuk, M. L., Sriwahyuningrum, A., Cahaya, I. M. E., Astuti, N. L. S., Setiawan, J., Tandirerung, W. Y., Rahmi, S., Nusantari, D. O., Indrawati, F., Fittriya, N. L., Aziza, N., Kurniawati, N., Wardhana, A., & Hasanah, T. (2023). *Metodologi penelitian*. Media Sains Indonesia: Bandung.