Proceedings of International Conference on Islamic Education (ICIED)
Impactful Education: Integrating Technology, and Ecotheological Values For a Sustainable Future
Volume 10, 2025 (p-ISSN 2477-3638 / e-ISSN 2613-9804)

Using Kappa Cohen for Expected Agreement of Math X-Quest Learning Media for in Support of SDGs-4 (Quality Education) Context

Putri Dwi Arianti[1], Taufiq Satria Mukti[2], Luthfiah Hamidah Nur'aini[3]

Correspondence:
tsatriam@uin-malang.ac.id

Affiliation:
Tadris Matematika, Faculty of Tarbiyah and Teacher Training, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia[1,3]
220108110038@student.uin-malang.ac.id

Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang[2]
taufiq.satria.2503419@students.um.ac.id

Abstract
Media quality needs to be proven in certain ways, one way is by expert assessment. In this paper, we examine the quality of Math X-Quest media which is assessed by two experts using the Cohen Kappa method. This activity aims to obtain validity in the form of feasibility of the contextual game-based Math X-Quest media that helps students understand the concept of three-variable linear equations through balancing activities, connecting abstract mathematical ideas with concrete experiences. The rater assessment consisted of aspects of the material and learning media. Analysis of the rater results was carried out using descriptive statistics with the Crosstabs technique. The results obtained showed that 2 raters had the same view on the quality of the assessment with a value of 0.69 spearman correlation. Meanwhile, the quality of X-Quest obtained a suitability score of 0.872 from the two expert assessors, indicating a near-perfect Cohen Kappa Agreement index. Thus, it can be concluded that X-Quest is suitable for use in mathematics learning which can support improving the quality of education as envisioned in SDG 4.

Keywords:
Kappa Cohen; Learning Media; Rater Expert; X-Quest

A. INTRODUCTION

Quality education is a key focus of the global Sustainable Development Goals (SDGs) campaign, particularly Goal 4, which emphasizes the importance of equity and improving the quality of education. Learning mathematics, as a discipline characterized by a high level of abstraction, requires a contextual approach to support students' understanding (Amalia et al., 2024; Reinke & Casto, 2022). Quality mathematics learning is a strategic effort to realize the SDGs (Olasoji et al., 2023). Students are trained to think logically and critically, and to connect abstract concepts with real-world phenomena in everyday life (Hana et al., 2024). Several efforts have been made to facilitate students' understanding of abstract concepts, including linking them to real-life facts and using teaching aids or learning media that can illustrate the material in a concrete and engaging way (Saharani & Abadi, 2024). Therefore, a contextual approach through learning media is crucial in supporting the movement for quality education that aligns with the SDGs.

In line with the importance of a contextual approach in mathematics learning, it is necessary to contextualize the learning media used by educators. This significantly supports the realization of quality education because learning media serves as a bridge between abstract contexts and students' concrete experiences (Abdulrahaman et al., 2020). Good learning media is not simply about displaying graphic-interactive elements or physical appeal. However, according to Farani (2013), Learning media must have good validity values related to content, design, language, and pedagogy to align with mathematics learning objectives. To ensure this quality, media validity assessment can be conducted through expert judgment. This assessment involves two or more experts who assess the media independently to provide more objective and

measurable results. This inter-expert assessment technique can be conducted using several statistical approaches, including the Content Validity Ratio (CVR) based on Lawshe's theory, Aiken's V, and Kappa theory (Aiken, 1985; Cohen, 1960; Zamanzadeh et al., 2015). These three techniques aim to assess the level of agreement or validity between experts on an instrument or media. However, in this study, the discussion will focus on Kappa theory, which not only measures agreement between experts but also takes into account the possibility of agreement that occurs by chance (expected agreement), so that the results are more accurate and objective (Sahu & Demirtas, 2024).

As a tangible manifestation of contextual learning, X-Quest was developed as an educational game that helps students understand the concept of Three Variable Linear Equation Systems (SPLTV) through balancing activities. This media is designed so that students can see and practice directly how the concept of equations works, so that learning is not only abstract but can also be experienced concretely through play and exploration activities. In this game, students not only memorize the procedures for solving SPLTV, but also experience deep mathematical thinking by placing "minerals" on the balance until they achieve a balance that represents the relationship between variables. Through this activity, students are trained to think logically, develop strategies, and make mathematical decisions based on evidence and direct observation. Thus, it can become a contextual media in line with SDG 4's goal of achieving quality and meaningful education.

In practice, expert assessments often lead to differing perspectives on the object being assessed (Ling, 2025). This is normal, considering that each expert has a different background, experience, understanding, and perspective in assessing learning media. Furthermore, unclear assessment criteria and varying interpretations of validity indicators can also influence assessment results (MacKay et al., 2017). This condition impacts the consistency of the object being assessed. Differences in perception between experts can be minimized by applying statistical analysis designed to measure the level of expert agreement, rather than simply calculating the average assessment score (Zec et al., 2017). One commonly used technique for this purpose is the Kappa technique. The Kappa coefficient is a statistical measure used to determine the level of agreement between experts on nominal-scale assessment categories.

This study refers to the findings put forward by Warrens (2015), who revealed that there are at least five ways to interpret the Kappa value depending on the context of its use: from the perspective of proportion of agreement, odds ratio, correlation coefficient, expected agreement, and error variance. Specifically, this study focuses on proving the expected agreement value that will be used in assessing learning media, thus producing a more objective and measurable assessment. This approach strengthens the reliability of the validation results because the Kappa method combines two or more expert assessment results and takes into account the possibility of random agreement. Furthermore, the Kappa method is flexible and can be applied to various fields, including education, psychology, health, and learning media research (Warrens, 2014).

Research using the Kappa method in the field of learning media assessment is still relatively rare. Until now, the application of Kappa in education has focused more on testing the reliability of instruments or questionnaires, while its application in the validation of learning media has been rare. Meanwhile, outside of education, the Kappa method has been widely applied in various disciplines to objectively and accurately measure the level of agreement between assessors. In psychology, research by Prawiro & Vereen (2024) Kappa method is used as a measuring tool to assess agreement between experts in analyzing quantitative data. In the health sector, the Kappa method is also used to assess the agreement between diagnoses between medical personnel and laboratories (McHugh, 2012). Research conducted Manik et al. (2025) in the clinical field, the Kappa method is applied to measure the reliability and consistency of the system in producing a diagnosis that is in accordance with the doctor's medical decision.

This research is expected to provide empirical evidence that the learning media development process has gone through a systematic and measurable validation process. The assessment is based not only on the average score from the experts but also on the level of agreement between experts, which is analyzed statistically using the Kappa method. This approach demonstrates that

the experts play an active role in ensuring the credibility, consistency, and reliability of the developed learning media. Therefore, it is hoped that the resulting media will be able to contribute to realizing quality education as stated in the fourth goal of the Sustainable Development Goals (SDGs).

B. METHODS

The use of the Cohen's Kappa method in this study aims to conduct a quantitative analysis of expert assessments of X-Quest as a mathematics learning media. This analysis was conducted in September 2025. The results of the Cohen's Kappa analysis will be used to draw conclusions about the feasibility of the X-Quest media.

The feasibility assessment process was carried out by two experts in the fields of mathematics education and educational media. The mathematics learning expert who assessed the X-Quest media was Dr. Syamsir Saenudin, M.Pd, while the mathematics learning media expert who assessed it was Dimas Femy Sasongko, M.Pd. The expert assessment was carried out on an ordinal scale of 1-5, with a higher score reflecting better quality. The expert assessment of X-Quest covered 20 items consisting of aspects of Material Quality (language and material suitability) and Media (design and ease of use). Data analysis was performed using descriptive statistics with the Crosstabs technique to obtain conclusions on the feasibility of the X-Quest media that had been developed from the perspective of the two raters' assessments.

C. RESULT & DISCUSSION

Cohen's Kappa analysis was used to estimate the agreement of expert assessments of an object. This study involved two experts, Dr. Syamsir Saenudin, M.Pd, an expert in mathematics learning, and Dimas Femy Sasongko, M.Pd, an expert in mathematics learning media. The assessment of the X-Quest media was carried out using an ordinal scale of 1-5, where 1 meant very poor, 2 meant poor, 3 meant fair, 4 meant good, and 5 meant very good. This assessment consisted of 20 assessment items covering Material Quality (language and material suitability) and Media aspects (design and ease of use). The rater assessment results obtained are presented in the following table:

Table 1. Expert assessment results (%)

Expert	Material		Media		
	Language	Relevance	Design	Language	Usability
Rater-1	84	84	76.67	80	90
Rater-2	84	84	76.67	90	100

Table 3. Results of Crosstabs analysis of 2 raters

Aspect	Result
Measure of Agreement Kappa	.872
Spearman Correlation	.069

The analysis results show that the agreement score from the raters' assessments is 0.872, while the Spearman correlation between the two is 0.69. The results of the Spearman analysis are used for testing compared to a significance value of 95%. If the Spearman value is > 0.05, then there is no significant difference between the assessments of the two experts. If the Spearman value is < 0.05, then the two raters have different views on the X-Quest media. Based on the above results, 0.069 > 0.05, so the raters have the same view on the quality of X-Quest.

Based on the Cohen's Kappa value of 0.872, conclusions can be drawn based on Landis & Koch (1977). The value obtained is 0.872, which, according to Table 4, falls into the category of almost perfect. This means that the X-Quest media meets the eligibility criteria as a learning media that can be used in learning to improve students' mathematical decision-making skills. The high level of agreement among experts indicates that the language, design, and material suitability components of X-Quest have met the principles of content validity. This means that this media has the potential to be widely used in contextual learning that supports the achievement of SDG-4.

Table 4. Cohen's Kappa Agreement Index

Kappa Statistic	Strength of Agreement
< 0.00	Poor
0.00 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost Perfect

The results of the Kappa analysis (κ = 0.872) indicate a high level of consistency among experts, which confirms that the X-Quest media meets the criteria for valid and reliable learning tools. This finding is highly relevant to Sustainable Development Goal 4, which aims to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all (Smith et al., 2020; United Nations, 2024). Learning media that is valid in terms of content, design, and pedagogy can directly support the realization of this goal because it promotes fairness, accessibility, and engagement in mathematics learning (February et al., 2025). In particular, media that bridges abstract mathematical contexts with students' concrete experiences is essential for developing conceptual understanding and meaningful learning (Arsyad et al., 2024; Widiana et al., 2020).

The integration of validated learning media into mathematics instruction aligns with global educational priorities emphasizing contextual and student-centered approaches. According to Sachdeva & Eggen (2021), quality mathematics learning should train students to think logically and critically while connecting abstract mathematical concepts to everyday experiences. Similarly, Prasetyowati (2020) emphasize that contextual media help learners visualize mathematical relationships and encourage reflective reasoning. Therefore, the X-Quest media, which transforms the abstract concept of three-variable linear equations into a balancing-scale activity, exemplifies the kind of pedagogical innovation envisioned in SDG 4 by making learning more inclusive, interactive, and relevant.

Beyond validating the quality of X-Quest, this study provides empirical evidence that measurable and statistically verified media contribute to sustainable educational improvement. Carvajal et al. (2025) highlights that achieving quality education requires continuous enhancement of teaching tools and digital innovations that are both context-responsive and evidence-based. Thus, the strong expert agreement obtained through the Kappa analysis not only strengthens the credibility of the media but also reflects a practical pathway to achieving SDG 4 targets through empirically validated instructional materials. Future implementation of X-Quest in classrooms can serve as a scalable model to promote mathematical literacy and sustainable learning outcomes.

D. CONCLUSION

Based on the results and discussion, it can be concluded that Kappa show the 0.872 is indicate almost perfect X-Quest Media category. This process know that Cohen Kappa can function to justify the feasibility of learning media conducted by experts. The involvement of at least two experts in the assessment of a particular object. Raters in determining media quality expectations must have the same view so that the distribution of assessments covers all ordinal scales used. Certainly, this provides additional insight in the field of education that the feasibility test process as a prototype can be carried out with empirical data proof. However, the researcher realizes that the two experts in this assessment is a limitation of the study. The researcher suggests that the assessment process needs to be further developed by involving more experts in accordance with the field so that the assessment of products in development goes through a more stringent stage to ensure their quality and usefulness.

ACKNOWLEDGMENTS

The publication of this article is part of a research activity supported by Recognition funding from FITK UIN Maulana Malik Ibrahim Malang.

REFERENCES

Abdulrahaman, M. D., Faruk, N., Oloyede, A. A., Surajudeen-Bakinde, N. T., Olawoyin, L. A., Mejabi, O. V, Imam-Fulani, Y. O., Fahm, A. O., & Azeez, A. L. (2020). Multimedia tools in the teaching and learning processes: A systematic review. *Heliyon*, *6*(11), e05312. https://doi.org/10.1016/j.heliyon.2020.e05312

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, *45*(1), 131–142.

Amalia, L., Makmuri, M., & Hakim, L. El. (2024). Learning Design: To Improve Mathematical Problem-Solving Skills Using a Contextual Approach. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, *7*(3), 2353–2366. https://doi.org/10.54371/jiip.v7i3.3455

Arsyad, M., Mujahiddin, M., & Syakhrani, A. W. (2024). The efficiency of using visual learning media in improving the understanding of science concepts in elementary school students. *Indonesian Journal of Education (INJOE)*, *4*(3), 775–787.

Carvajal, A. L. P., Fernandez, T. M., Pangilinan, A. M., Obod, M. M., Amihan, S. R., Sanchez, R. D., Sanchez, A. M. P., & Sanchez, J. J. D. (2025). Future-Proofing Teachers in Reframing Teacher Education Curriculum in the Philippines: Basis for Policy Recommendations. *International Journal of Open-Access, Interdisciplinary and New Educational Discoveries of ETCOR Educational Research Center (IJOINED ETCOR)*, *4*(2), 235–252.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Farani, R. (2013). The Evaluation Of Instructional Multimedia Product For English Learning. *Journal of English and Education (JEE)*.

February, P., Alter, G., Buendgens-Kosten, J., Müller, F. J., & Di Paolo, A. (2025). (DIGITAL) MEDIA AND MATERIALS FOR LEARNING. In *All means all! -OpenTextbook for diversity in education* (pp. 1115–1148). All-means-all.

Hana, A. Y., Sa'diyah, H., Nahar, Z. K., & Fakhriyana, D. (2024). Pendidikan Berkualitas untuk Mewujudkan Sustainable Development Goals (SDGs) Melalui Pembelajaran Matematika: Systematic Literature Review. *ARITHMETIC: Academic Journal of Math*, *6*(1), 79–100.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Ling, J. H. (2025). A Review of Rubrics in Education: Potential and Challenges. *Indonesian Journal of Innovative Teaching and Learning*, *2*(1), 1–14. https://doi.org/10.64420/ijitl.v2i1.197

MacKay, R. S., Kenna, R., Low, R. J., & Parker, S. (2017). Calibration with confidence: a principled method for panel assessment. *Royal Society Open Science*, *4*(2), 160760.

Manik, I. M. S. B., Rizal, C., & Kurniawan, H. (2025). ANALISIS AKURASI DAN KEANDALAN METODE CERTAINTY FACTOR PADA SISTEM PAKAR DIAGNOSTIK MEDIS. *Jurnal Multimedia Dan Teknologi Informasi (Jatilima)*, *7*(03), 702–707.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Olasoji, O. V., Ugwu, U. F., & Onoh, D. O. (2023). The Role of Mathematics Education in Achieving Sustainable Development Goals (SDGs). *ESUT Journal of Education*, *6*(1), 247–255.

Prasetyowati, D. (2020). Development of Learning Videos for Junior High School Math Subject to Enhance Mathematical Reasoning. *International Journal of Education and Practice*, *8*(1), 18–25.

Prawiro, F., & Vereen, A. (2024). REPRESENTASI SOSIAL TENTANG GANGGUAN JIWA PADA MASYARAKAT YANG INGIN MENCARI BANTUAN PSIKOLOGIS FORMAL (STUDI PADA MASYARAKAT BANTEN YANG RELATIF MODERN). *MANASA*, *13*(2), 74 - 91. https://doi.org/10.25170/manasa.v13i2.6099

Reinke, L. T., & Casto, A. R. (2022). Motivators or conceptual foundation? Investigating the development of teachers' conceptions of contextual problems. *Mathematics Education Research Journal*, *34*(1), 113–137.

Sachdeva, S., & Eggen, P.-O. (2021). Learners' Critical Thinking About Learning Mathematics. *International Electronic Journal of Mathematics Education*, *16*(3), em0644. https://doi.org/10.29333/iejme/11003

Saharani, T., & Abadi, A. M. (2024). The Development of Interactive Learning Media Based on Discovery Learning Oriented to Students' Concept Understanding and Mathematical Disposition. *AL-ISHLAH: Jurnal Pendidikan*, *16*(4). https://doi.org/10.35445/alishlah.v16i4.5681

Sahu, S., & Demirtas, H. (2024). A Linear Relationship between Correlation and Cohen's Kappa for Binary Data and Simulating Multivariate Nominal and Ordinal Data with Specified Kappa Matrix. *ArXiv Preprint ArXiv:2404.14149*.

Smith, N. M., Hoal, K. E. O., & Thompson, J. F. H. (2020). Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all. In *Mining, materials, and the sustainable development goals (SDGs)* (pp. 29–38). CRC Press.

United Nations. (2024). *Sustainable Development Goal 4: Quality Education*. https://www.un.org/sustainabledevelopment/education

Warrens, M. J. (2014). New Interpretations of Cohen's Kappa. *Journal of Mathematics*, *2014*, 1–9. https://doi.org/10.1155/2014/203907

Warrens, M. J. (2015). Five ways to look at Cohen's kappa. *Journal of Psychology & Psychotherapy*, *05*(04). https://doi.org/10.4172/2161-0487.1000197

Widiana, I. W., Tegeh, I., & Parwata, I. (2020). Improving Students' Factual Knowledge with Concrete Media through Observing Activities in Scientific Approaches in Elementary Schools. *Journal of Education and E-Learning Research*, *7*(3), 293–299.

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Journal of Caring Sciences*, *4*(2), 165–178. https://doi.org/10.15171/jcs.2015.017

Zec, S., Soriani, N., Comoretto, R., & Baldi, I. (2017). High Agreement and High Prevalence: The Paradox of Cohen's Kappa. *The Open Nursing Journal*, *11*(1), 211–218. https://doi.org/10.2174/1874434601711010211