EDUVELOP
JOURNAL OF ENGLISH EDUCATION AND DEVELOPMENT
Nationally Accredited Journal Decree No.10/C/C3/DT.05.00/E5/2025 (Sinta 2)

Volume 9, No. 1, March 2026 • **Page: 241 - 252**
DOI: https://doi.org/v9i1.6093
ISSN 2597-713X (print)
ISSN 2597-7148 (online)

OPEN ACCESS: Research Article

Feedback as a Determinant of Online Assessment Validity: A Multidimensional Analysis in Indonesian EFL Higher Education

Mursalim[1*], Alberth[2], La Ode Nggawu[3], Anugrah Puspita Ayu[4], Wahyudin Madil[5]
[1]Universitas Halu Oleo, Kendari, Indonesia

*Correspondence e-mail: mursalim@uho.ac.id

Received : 2025-12-19	**Accepted :** 2025-12-19
Revision : 2026-02-13	**Published :** 2026-03-31

Abstract
Although LMS-based online assessment is widely used, there is a paucity of empirical research studies that have investigated its multidimensional validity from students' perspectives, especially in Indonesian higher education. The proposed cross-sectional survey research examines students' perceptions of the validity and quality of online assessment in the English Education Department at Universitas Halu Oleo (UHO) delivered through the E-Green SPADA (Moodle-based) course platform. The study conceptualizes perceived validity based on multidimensional and argument-based validity frameworks that are prominent in contemporary research. A 25-item questionnaire (5-point Likert scale) that included construct validity, assessment fairness, feedback & support, learning impact, and affective impact was administered to a stratified random sample of 120 students during the final week of the semester. Exploratory factor analysis showed a five-factor structure (KMO = 0.89; Bartlett $\chi^2(300) = 1650.3$, $p < .001$), which explained 68.2% of the total variance, with acceptable internal consistency ($\alpha = .76–.89$). The descriptive results showed that construct validity ($M = 4.21$, $SD = 0.61$) and assessment fairness ($M = 4.08$, $SD = 0.65$) were highly rated, while feedback & support received relatively low scores ($M = 3.49$, $SD = .79$). Feedback & support was the strongest predictor of overall perceived validity ($\beta = .458$, $p < .001$) in multiple regression analysis, and the full model accounted for 62.7% of the variance ($R^2 = .627$). These findings provide support for the consequential facet of validity in that the prompt, comprehensive, and accessible feedback can influence students' confidence in online assessment. The research adds evidence from the Indonesian EFL higher education context and highlights the importance of LMS design features, instructor training, and feedback literacy programs that facilitate dialogic and responsive feedback practices.

Keywords: *Online Assessment Validity; Student Perceptions; Feedback And Auspport; Learning Management System; Exploratory Factor Analysis*

DOI: https://doi.org/v9i1.6093

1. Introduction

The fast move to online education all over the world has affected assessment practices in higher learning. Computer-based testing and LMS-based assessment have become more adaptable, scalable, and convenient formats of technology-mediated assessments (Adedoyin and Soykan, 2023). Meanwhile, these innovations have led to queries regarding the merits, equity, and pedagogical integrity of assessment especially when efficiency and convenience are given precedence over real evaluation of learning. Validity in language education is not just a statistical measure; tests should be able to represent the skills of learners to communicate in real-life situations instead of just isolated linguistic knowledge (Mate et al., 2021; Norris and Davis, 2021).

Under the Indonesian higher education scene, national policy frameworks such as the Undang-Undang No. 12 Tahun 2012 tentang Pendidikan Tinggi have provided that the credibility, transparency, and accountability of assessment practices in higher education institutions are required. With the growing adoption of LMS including E-Green SPADA in institutions, it is vital to show the legitimacy of online assessment. Nonetheless, the implementation of the principles of validity in the digital context is not always balanced, limited by the infrastructure, differences between the digital competence of the instructors, and inconsistent feedback mechanisms. The research studies carried out during the COVID-19 era also demonstrate that there are still connectivity issues, imbalanced device accessibility, and delays in feedback delivery, all of which contribute to how students view the quality of assessments and perceive them as fair. These contextual factors do not just influence the nature of learning conditions, but provide a dilution of interpretive transparency and the perceived validity of the assessment results.

Contemporary validity theory emphasizes that validity is best understood as an argument-based, multidimensional construct (Kane, 2016; Hill et al., 2022). In this framework, assessment validity is supported by a chain of inferences linking observed performance to interpretation and use, encompassing scoring, generalization, extrapolation, and implications. Interpretive inferences concern whether observed performances are meaningfully interpreted as indicators of the intended construct, while extrapolation inferences address whether those interpretations extend beyond the specific assessment task to broader domains of competence. This study operationalizes perceived validity across five dimensions that correspond to these inferential links. Construct Validity reflects the extrapolation inference, measuring alignment between tasks and intended language abilities. Assessment Fairness maps onto the generalization inference, addressing consistency and equity across learners. Feedback & Support and Learning Impact relate primarily to the implications inference, capturing how assessment information is interpreted and applied to support learning. Finally, Affective Impact reflects students' confidence and trust, which influence the interpretive acceptance and use of assessment outcomes.

Specifically, feedback is the key to numerous inferences and cannot be viewed merely as part of consequential validity. The studies concerning feedback literacy emphasize the fact that students should have the skills to interpret, judge, and respond to feedback instead of passively taking it. The discipline-focused model offered by Naomi Winstone

DOI: https://doi.org/v9i1.6093

and David Carless (2019) highlights that successful feedback involves the need to cultivate evaluative judgment, whereas the same study (2023) also reveals the dynamic nature of the feedback literacy between teachers and students. In cases where instructors are not able to offer clear, actionable, and construct-relevant feedback, the capacity of the students to obtain a valid meaning of the results of assessment is reduced.

Within an argument-based validity framework, feedback mediates both interpretive and extrapolation assumptions. First, clear and criteria-referenced feedback strengthens interpretive inferences by clarifying how scores relate to construct representation, thereby enhancing students' understanding of what their performance demonstrates. Second, feedback that supports revision, transfer, and future application contributes to extrapolation inferences, reinforcing the claim that assessment performance reflects enduring communicative competence rather than task-specific behavior. In LMS-mediated environments, delayed, vague, or technologically inaccessible feedback can disrupt this inferential chain, weakening students' trust and reducing the plausibility of validity claims (Carless et al., 2021; Malecka et al., 2022; Winstone & Carless, 2019).

By explicitly situating feedback within Kane's inferential chain, this study moves beyond a descriptive account of student perceptions and advances a conceptually grounded contribution to digital assessment validity discourse. It demonstrates how feedback practices, fairness considerations, and affective engagement function not merely as peripheral consequences of assessment, but as integral components that sustain the interpretive coherence and extrapolatory strength of LMS-based assessment in Indonesian higher education.

Although previous literature has investigated online assessment validity in general terms, relatively few studies have empirically connected feedback literacy frameworks with multidimensional validity constructs in EFL contexts. This study addresses that gap by incorporating feedback quality as a predictive dimension of perceived validity, situating it within an argument-based framework of interpretive, extrapolative, and implicative inferences.

The findings carry concrete implications for instructional and institutional practice. For example, LMS-based computerized formative quizzes with adaptive, criteria-referenced feedback can enhance interpretive clarity by explicitly linking performance outcomes to construct definitions, thereby strengthening the evidentiary basis for score meaning (Kane, 2016). Peer feedback calibration sessions – where students evaluate sample responses using shared rubrics and discuss discrepancies – can foster evaluative judgment, a core component of feedback literacy (Winstone & Carless, 2019), and reinforce assumptions of fairness and generalization. Additionally, audio or video comments explicitly aligned with rubric descriptors may support dialogic engagement and actionable interpretation, addressing concerns about feedback usability and timeliness in digital environments (Carless et al., 2021; Malecka et al., 2022).

For institutional stakeholders, investing in structured feedback literacy development for both instructors and students – including professional development on rubric alignment, dialogic digital tools, and construct-referenced feedback design – can enhance the interpretive coherence and extrapolative strength of LMS-mediated

DOI: https://doi.org/v9i1.6093

assessments. In doing so, universities not only meet accountability expectations embedded in national quality assurance frameworks but also reinforce the multidimensional validity argument underpinning digital assessment practices in Indonesian higher education.

This study was guided by the following research questions:

1. What dimensions best represent students' perceptions of online assessment validity in LMS-based English education courses?
2. Which dimensions of online assessment validity, particularly feedback quality, most significantly predict students' overall perceptions of validity in LMS-based English education courses?
3. Are there significant differences in students' perceptions of assessment validity across course types or academic cohorts in the E-Green SPADA system?

2. Methods

2.1 Research Design

This study employed a quantitative, cross-sectional survey design to systematically capture students' perceptions of the validity of online assessments administered through the E-Green SPADA learning management system at the Faculty of Teacher Training and Education (FKIP), Universitas Halu Oleo (UHO) in Indonesia. A cross-sectional approach was considered appropriate because it enabled the simultaneous examination of multiple validity dimensions at a specific time point following students' completion of online course assessments. The quantitative approach ensured objectivity, cohort comparability, and the use of measurable indicators suitable for statistical modeling (Creswell, 2018).

2.2 Population and Sampling Procedures

The target population consisted of 645 undergraduate students who were enrolled in the English Education Department in the 2024-2025 academic year. To maintain the representativeness of the sample, a stratified random sampling process was applied to the sample variables, which included an individual student cohort (2021-2024 intakes) and an individual course type (Web-Based Language Learning, Computational Linguistics, Computer-Assisted Language Learning, Technology-Enhanced Language Learning and Business English).

One hundred and twenty students (18.6% of the population) were proportionally selected based on stratification and the distribution of enrollment. This sample size is adequate, meets the minimum of finite populations (Krejcie and Morgan, 1970), and corresponds to the factor analytic rules of 5-10 respondents per item (Hair Jr et al., 2020). The participants were from all the academic years and both sexes.

The inclusion criteria were as follows: (a) the participants had to have completed at least one full E-Green SPADA online course, (b) at least one graded online assessment, (c) had to sign the voluntary consent. The involvement was anonymous and was not connected to academic status.

DOI: https://doi.org/v9i1.6093

2.3 Instrument Development and Validation

A questionnaire of 25 items was created in order to assess the validity of online assessments for students. The instrument was theoretically based on the multidimensional argument-based validity models (Kane, 2016; Hill et al., 2022; Mate et al., 2021) and operationalized along five dimensions:

1. Construct Validity (5 items)
2. Assessment Fairness (5 items)
3. Feedback & Support (5 items)
4. Learning Impact (5 items)
5. Affective Impact (5 items)

It used a five-point Likert scale (1 = Strongly disagree to 5 = Strongly agree).

2.4 Content Validation

The initial item pool of 30 items was based on the previous assessment validity and feedback literacy frameworks. Items on the questionnaire were tested on their clarity, cultural appropriateness, and construct alignment by three language assessment and educational measurement professionals. According to the experts' comments, the final 25-item instrument was created after as many redundant and ambiguous items as possible were eliminated.

2.4.1 Pilot Testing

The questionnaire was piloted among 30 students who are not in the sample. Cosmetic changes were made in terms of wording to increase clarity. The alpha coefficients of Cronbach were between .82 and.89 which indicated high internal consistency.

2.5 Data Collection Procedures

Data collection occurred in the last 2 weeks of the first semester (May-June 2025). The questionnaire was conducted through the E-green SPADA system via Google Forms. The information provided to the subjects included details concerning the goal and purpose of the study, confidentiality and the voluntary nature of participation. The mean time was approximated to be 10-12 minutes.

2.6 Data Analysis Procedures

The entire analysis was done using the IBM SPSS Statistics version 27. After the screening of the data based on completeness and normality, descriptive statistics (means, standard deviations, skew, and kurtosis) were calculated. The values of skewness and kurtosis were also at reasonable levels (+-2), which imply that there was approximate univariate normality.

2.7 Exploratory Factor Analysis (EFA)

Principal Axis Factoring with Promax (oblique) rotation was done to investigate the latent structure of the instrument. The sampling adequacy was ensured (KMO = 0.89), and the Test of Sphericity $\chi^2(300) = 1650.3$, $p < .001$ was also significant, which means that the factor analysis is appropriate. Retention of factors was done based on eigenvalues more than 1 and scree plot. Items with loadings ≥ .40 were retained.

DOI: https://doi.org/v9i1.6093

3. Results

3.1 RQ1: Dimensions of perceived validity

Exploratory factor analysis proved to support a five-factor structure as follows in terms of multidimensionality of online assessment validity (KMO = 0.89; Bartlett $\chi^2(300) = 1650.3$, $p < .001$), which accounted for 68.2 percent of total variance. The five measures - Construct Validity, Assessment Fairness, Feedback and Support, Learning Impact, and Affective Impact displayed good internal consistency (a =.76-.89).

Descriptive outcomes showed that Construct Validity (M = 4.21, SD = 0.61) and Assessment Fairness (M = 4.08, SD = 0.65) were the highest rated meaning that the assessments were perceived to be well-purposed in terms of learning goals and consistent in their administration. Feedback and Support were lower (M = 3.49, SD =.79) with the perceived constraints in timeliness, specificity and access. There was a moderate rating in Learning Impact (M = 3.77, SD = 0.70) and Affective Impact (M = 3.88, SD = 0.68), which indicated that assessments had a moderate effect on motivation and confidence.

In the interpretive model parallel to Kane's framework, Construct Validity refers to the extrapolation inference, evaluating consistency with the real-life language skills; Assessment Fairness to the generalization inference, evaluating fairness on the score assigned to students; Feedback & Support, Learning Impact, and Affective Impact are mainly associated with the implications inference, which affects how students receive and use assessment outcomes.

3.2 RQ2: Predictors of overall perceived validity

Multiple regression analysis identified Feedback & Support as the strongest predictor of overall perceived validity ($\beta = .458$, $p < .001$), followed by Learning Impact ($\beta = .206$, $p = .007$) and Construct Validity ($\beta = .184$, $p = .022$). Assessment Fairness and Affective Impact contributed less once other variables were accounted for. The full model explained 62.7% of the variance ($R^2 = .627$), highlighting that while structural and affective factors are important, the quality of feedback plays a central role in shaping students' trust in online assessment.

Theoretical interpretation: Feedback operates across multiple inferences in Kane's chain. It informs students about their scores (scoring inference), clarifies the generalizability of performance (generalization inference), supports extrapolation to real-world language use, and mediates consequential outcomes such as learning engagement and trust

3.3 RQ3: Group differences

An analysis of variance was found to be significant ($F(5, 114) = 3.76$, $p = .004$, $\eta^2 = .14$) between course types. Comparisons that were made post hoc (Tukey HSD) revealed that students in Business English had a stronger perceived validity (M = 4.15) than students in Web-Based Language Learning (M = 3.68) and Computational Linguistics (M = 3.71). Generally, no significant difference was found in other courses.

There were no significant differences in cohort years ($F(3, 116) = 1.26$, $p = .293$, $\eta^2 = .03$), which indicated that the perceptions did not differ in accordance with academic levels. These trends indicate that course design based on the authenticity of the tasks

DOI: https://doi.org/v9i1.6093
and the alignment affect the perceptions of the students as opposed to academic seniority.

The results support an empirical argument-based conceptualization of multidimensional LMS-mediated assessment validity. Although other structural variables, such as Construct Validity and Assessment Fairness, give the underlying feeling of confidence, Feedback and Support come out as the most important contributor of perceived validity, thus its centrality in the inferential chain of Kane. Learning Impact and Affective Impact reinforce this and stress that the involvement, stimulation, and trust of students form the basis of the holistic validity of online evaluations.

4. Discussion

4.1 RQ1: Dimensions of perceived validity

The research has found that there are five dimensions of Construct Validity, Assessment Fairness, Feedback and Support, Learning Impact, and Affective Impact, which cumulatively explain the results for students on the online assessment validity in English courses using LMS. The construct validity and Assessment Fairness scores are high thus showing that students believe that the assessments are in line with the learning objectives and are conducted in a fair manner. Construct Validity, according to Kane, is more of a reflection of the extrapolation inference, which connects the performance of the students on the LMS-based activities with the use of language in the real world, whereas Assessment Fairness is more of the generalization inference, which involves consistency of the performance judgments of the learners.

On the other hand, the lower score of Feedback and Support highlights that students feel that the feedback is not timely, not specific, and unavailable which hinders their capacity to interpret the assessment results and use them as a guide to learning. Feedback makes various inferences: feedback informs the scoring inference (by clarifying scores), the extrapolation inference (by putting the performance in context), and the implications inference (by directing the further actions of the learners). These results underpin the key role of feedback in developing the trust and perception of assessment credibility in students, which posit that structural soundness is not a sufficient condition toward the holistic validity.

4.2 RQ2: Predictors of overall perceived validity

Regression analyses revealed that Feedback and Support exert the most predictive power over the overall perceived validity followed by Learning Impact and Construct Validity. This means that the belief that students have in the assessment is based on pedagogical interaction but not the structural design per se. The impact of learning and the Affective Impact are also involved, and the connection between the implications inference and motivation, engagement, and emotional response is established. Practically, this implies that LMS-based assessments are not merely psychometrical in nature, but are held as being valid when the students are provided with actionable, timely and dialogic feedback.

DOI: https://doi.org/v9i1.6093

These findings are consistent with the previous studies that focus on feedback literacy (Winstone and Carless, 2019, 2023; Carless et al., 2021). Students are better placed to interpret scores appropriately, generalize learning, and apply the knowledge to real-life language tasks when the instructors are ready to give such contextually meaningful feedback. Feedback, therefore, serves as a key element that ties the scoring, generalization, extrapolation and consequential inferences together supporting the argument-based validity chain.

4.3 RQ3: Group differences

A high level of variation between the types of courses implies that the perception of validity depends on the correspondence between tasks and the use of authentic language. Business students noted an increased level of validity, probably because of the real-life approach in course testing. Cohort-year comparisons revealed no significant difference, which means that there is no variation in the perceptions concerning the levels of studies. Such results indicate that the effectiveness of LMS-based assessments in the context of supporting the chain of inferences between scoring and implications is moderated by contextual factors, including the authenticity of the task and the design of the course.

4.5 Theoretical and Practical Implications

The research adds to the discourse of validity in the field of digital assessment because it shows how the argument-based framework presented by Kane can be applied to understand the way in which students perceive the assessment. The feedback appears to be the key process that connects various inferences: it confirms scores, the applicability of learning in general, and informs the further actions. Thus, feedback literacy, which must support a timely, actionable, and participatory feedback, should be of primary concern of LMS design, instructor training, and institutional policies.

The findings also demonstrate that the combination of pedagogical and affective aspects into the design of online assessment is essential. Credibility requires construct alignment and fairness which should be supported by high quality feedback and sensitivity to emotional and motivational reactions of students so that completely valid assessment results can be attained.

5. Conclusion

Although in this study strong evidence of the multidimensional validity of LMS-based assessments was presented, one should admit that there are a number of limitations. First, there could be common method bias since the study will be based on self-reported survey data that will be collected at one time, and thus will inflate observed relationships. The triangulation of student perceptions with instructor assessments or with LMS analytics might be used in future studies to enhance the validity of the findings. Second, the research was carried out in one institution, Universitas Halu Oleo, which can be a limitation to the external validity of the findings. Multi-institutional research would help to explain the wider applicability of these insights to other contexts of LMS implementation and courses. Third, the cross-sectional design only captures perceptions at a given time making causal inference impossible. Longitudinal or mixed-

DOI: https://doi.org/v9i1.6093

method studies would help answer the question of how the feedback practices change and affect the validity chain with time. Fourth, the feedback was considered as a unitary construct; the disaggregation of the dimensions like timeliness, specificity, interactivity and modality would contribute to revealing which of the elements have the strongest impact on students developing trust and learning outcomes. Lastly, the overall moderators such as digital literacy, class size and authenticity of the task can influence perceived validity and should be investigated further.

Although limited, the study proves that the online assessment in LMS-based systems, which is practiced in Indonesian higher education, is perceived to be generally valid, fair, and meaningful. The paper correlates the perceptions of students with the argument-based chain of validity of Kane and demonstrates that Construct Validity (extrapolation) and Assessment Fairness (generalization) are used to generate structural credibility, and that Feedback and Support, Learning impact and Affective impact are used to create scoring interpretation, extrapolation and consequential inferences.

The key of these dimensions is Feedback & Support as it is the primary determinant of perceived validity and mediates various inferences: it helps clarify the scores, contextualize performance to apply it into practice, and influence the future actions of the learners. The best, timely and dialogic feedback increases the accuracy of the interpretations, builds on the reliability of the assessment results and promotes the transferability of the learning process making it the most essential attribute in the holistic perceptions of the validity. Applied implications are:

1. Improving LMS to facilitate the use of interactive and actionable feedback.
2. Encouraging instructors and students to be feedback literate in order to maximize interpretive and consequential use.
3. Building congruence between assessment design and authentic tasks and institutional quality assurance practices that can help sustain structural and pedagogical validity.

Future studies are advised to examine the longitudinal impact of feedback, at the same time, disaggregate the feedback dimension and study contextual moderators, including digital literacy, class size, and the authenticity of the task. This study proposes a theoretically based framework of designing, assessing, and refining LMS-based testing in tertiary learning institutions by a clearly stated association of student perceptions to the chain of inferences proposed by Kane, as a guideline toward better student learning, institutional responsibility, and the overall validity of online assessment systems.

Conflicts of Interest

Authors declare no conflict of interest.

References

Almulla, M. (2020). Developing 21st-century competencies through project-based learning. *International Journal of Instruction*, 13(1), 233–248.

Al-Tamimi, A., & Lin, S. (2018). Speaking anxiety among university students and its impact on performance. *International Journal of Education and Literacy Studies*, 6(1), 39–48.

Alqurashi, E. (2022). What do students engage with the most? A comparative study between high and low achieving students within online learning environments. *Open Learn.*37, 219–234. doi: 10.1080/02680513.2020.1758052

Ambawani, S., Astasari, I., & Rukiati, E. (2025). Overcoming barriers to EFL speaking proficiency: A multidimensional analysis of language learning challenges. *Journal of English in Academic and Professional Communication JEAPCO*, *11*(1), 2025. https://doi.org/10.25047/jeapco.v11i1.5769

Aubrey S. (2022). *Enhancing long-term learner engagement through project-based learning*

Bell, S. (2015). PJBL as a tool for deeper learning in the 21st century. *The Clearing House*, 83(2), 39–43.

Birdman J., Wiek A., Lang D. J. (2022). Developing key competencies in sustainability through project-based learning in graduate sustainability programs. *Int. J. Sustain. High. Educ.*23, 1139–1157. 10.1108/IJSHE-12-2020-0506

Blumenfeld P. C., Soloway E., Marx R. W., Krajcik J. S., Guzdial M., Palincsar A. (1991). Motivating project-based learning: sustaining the doing, supporting the learning. *Educ. Psychol.*26, 369–398. 10.1080/00461520.1991.9653139

Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity achievement emotions and academic performance: A meta-analysis. *Educational psychology review*, *33*(3), 1051-1095.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches (4th ed.)*. SAGE Publications.

Derakhshan, A., & Karimian, Z. (2021). Exploring the relationship between speaking skills and academic performance. *Language Teaching Research Quarterly*, 21, 45–59.

Dehbozorgi, N. & Kunuku, M. T. (2023). Exploring the influence of emotional states in peer interactions on students' academic performance. *IEEE Trans. Educ.*67, 405–412. doi: 10.1109/te.2023.3335171

Dewaele, J.-M., & Li, C. (2020). Emotions in second language acquisition: A critical review and research agenda. *Foreign Language Annals*, 53(2), 343–362.

Du X., Lundberg A., Ayari M. A., Naji K. K., Hawari A. (2022). Examining engineering students' perceptions of learner agency enactment in problem- and project-based learning using Q methodology. *J. Eng. Educ.*111, 111–136. 10.1002/jee.20430

Elov, B., Abdullayeva, I., Raupova, L., Kholikov, A., & Mirkasimova, M. (2025). Speaking exams with less anxiety in Intelligent Computer-Assisted Language Assessment (ICALA): mirroring EFL learners' foreign language anxiety, shyness, autonomy, and enjoyment. *Language Testing in Asia*, *15*(1), 3.

Ferecbeyli, N. (2025). Project-Based Learning in Language Teaching. *EuroGlobal Journal of Linguistics and Language Education*, *2*(5), 107–118. https://doi.org/10.69760/egjlle.2505009

Ferrero, M., Vadillo, M. A., & León, S. P. (2021). Is project-based learning effective among kindergarten and elementary students? A systematic review. *PloS one*, *16*(4), e0249627.

Gabuardi V. M. F. (2021). Project-based learning: Boosting 21st century skills. *Revista estudios*,

43, 340–419. doi: 10.15517/re.v0i43.49335
Gökçen, A., & Güngör, S. (2023). How project-based learning promotes emotional involvement in university students. *Journal of Language and Education*, 9(2), 22–30.
Habbal F., Kolmos A., Hadgraft R. G., Holgaard J. E., Reda K. (2024). "*Problem and project-based learning at aalborg," in Reshaping Engineering Education* (Springer Nature Singapore), 199–222
Hadibarata, T., Hidayat, T., & Kwabena, J. (2023). Problem based learning in engineering course in Malaysia. *Acta Pedagogia Asiana*, *2*(2), 95-105.
Hidayati, D., Novianti, H., Khansa, M., Slamet, J., & Suryati, N. (2023). Effectiveness project-based learning in ESP class: Viewed from Indonesian students' learning outcomes. *International Journal of Information and Education Technology*, *13*(3), 558-565.
Huang, L., & Liu, M. (2025) The impact of technology on foreign language anxiety: A systematic review of empirical studies from 2004 to 2024. *Humanit Soc Sci Commun* 12, 1665. https://doi.org/10.1057/s41599-025-05921-6
Kim, Y. (2019). Emotional engagement in ESL speaking through project-based tasks. *Asian EFL Journal*, 21(3), 156–175.
Lestari, R. D., Syahid, A., & Widiastuty, H. (2025). *Strategies for Managing EFL Speaking Anxiety in an Indonesian University: A Sequential Explanatory Study*. https://doi.org/10.25217/jed.v3i01.5645
María, V., & Gabuardi, F. (n.d.). Project-Based Learning: boosting 21st century skills. *Revista Estudios*, (43), 2021. Retrieved https://orcid.org/0000-0002-8401-122X
Menon, M. E., Markadjis, E., Theodoropoulos, N., & Socratous, M. (2017). Higher education choices: What drives students' decisions? *Studies in Higher Education*, 42(2), 252–267.
Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation (4th ed.)*. Jossey-Bass.
Meyer, D. K., & Turner, J. C. (2019). Emotional engagement and motivation to learn: A review of current research. *Educational Psychology Review*, 31(2), 399–419.
Mims, C. (2020). Meaningful learning and student participation in higher education. *Journal of Instructional Pedagogies*, 25, 112–126.
Mota, F. B., Cabral, B. P., Braga, L. A. M., & Lopes, R. M. (2025). Mapping the global research on project-based learning: a bibliometric and network analysis (2014–2024). In *Frontiers in Education* (Vol. 10). Frontiers Media SA. https://doi.org/10.3389/feduc.2025.1522694
Nifriza, I., Rahmalia, I., & Reswary, D. A. (2025). The Effect of Project-based Learning on the Improvement of Listening Skills in Junior High School Students. *Linguistic, English Education and Art (LEEA) Journal*, *8*(3), 305–310. https://doi.org/10.31539/zrmxnw51
Nurmala, I., Bukhari, D., & Habibi, M. A. (2021). Project-Based Learning (PJBL) to Improve Students' Speaking Skills in EFL Classroom. *Journal of Language Teaching and Research*, 12(4), 682–688. https://doi.org/10.17507/jltr.1204.06.
Pekrun, R., Goetz, T., & Perry, R. P. (2017). *Understanding learners' emotions through control-value theory*. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), Handbook of emotions in

education (2nd ed., pp. 120–141). Routledge.

Rusyda, R. N., & Riyanto, S. (2023). Enhancing EFL Students' Speaking Performance through Digital Project-Based Learning. *Journal of English Language Education and Literature*, 7(1), 45–58. https://doi.org/10.31227/osf.io/3h2xf.

Sa'adah, H., & Wahyuni, S. (2024). A critical review: Students' outcomes of project based Learning (PjBL). *Prominent: Journal of English Studies*, *7*(2), 1-15.

Sari, F. M., & Oktaviani, L. (2020). The Use of Project-Based Learning in Improving Students' Speaking Ability and Creativity. *Journal of English Language Teaching and Linguistics*, 5(3), 349–359. https://doi.org/10.21462/jeltl.v5i3.418.

Siahaan, B. L., & Siahaan, M. M. (2023). The Implementation of Project Based Learning Connected Wwth Digital Technology to Increase Students Speaking Competence of Madrasah Aliyah Negeri (MAN) Simalungun Regency. *AL-ISHLAH: Jurnal Pendidikan*, *15*(1), 497–506. https://doi.org/10.35445/alishlah.v15i1.2580

Solhi, M. (2024). The impact of EFL learners' negative emotional orientations on (un) willingness to communicate in in-person and online L2 learning contexts. *Journal of Psycholinguistic Research*, *53*(2), 26.

Thomas, J. W. (2020). *A review of research on project-based learning*. Buck Institute for Education.

Thomas, N. J., Baral, R. (2023) Mechanism of gamification: Role of flow in the behavioral and emotional pathways of engagement in management education, *The International Journal of Management Education*, 21, 1, https://doi.org/10.1016/j.ijme.2022.100718.

Tran, N., Hoang, D. T. N., Gillespie, R., Yen, T. T. H., & Phung, H. (2024). Enhancing EFL learners' speaking and listening skills through authentic online conversations with video conferencing tools. *Innovation in Language Learning and Teaching*, 1-11.

Wijaya, A. N., Al Faruq Hasyim, U. A., Zuniati, M., & Khatan Smita, M. (2025). The Effectiveness of Project-Based Learning Toward Students' Speaking Performance and Skill. *Attractive: Innovative Education Journal*, *8*(1). https://www.attractivejournal.com/index.php/aj/

Zein, S., Sukyadi, D., Hamied, F. A., & Lengkanawati, N. S. (2020). English language education in Indonesia: A review of research (2011–2019). *Language Teaching*, *53*(4), 491-523.

Zhong, J., Ismail, L., & Lin, Y. (2025). Investigating EFL students' engagement in project-based speaking activities: from a multi-dimensional perspective. *Frontiers in Psychology*, *16*. https://doi.org/10.3389/fpsyg.2025.1598513