

An Evaluation Model to Improve Student Learning Effectiveness in Learning Outcome Evaluation

Abdul Hasan Saragih¹, Harun Sitompul², and Muslim³

{ abdulhasan@unimed.ac.id¹, profharun@unimed.ac.id², muslim@unimed.ac.id³ }

Department of Mechanical Engineering Education, Faculty of Engineering,
Universitas Negeri Medan, Medan, North Sumatra, Indonesia ^{1,2,3}

Abstract. This study developed a valid, practical, and effective learning evaluation model to enhance student learning effectiveness in the Learning Outcome Evaluation course of the Mechanical Engineering Education Study Program. Using a Research and Development (R&D) method with a modified Borg & Gall model (eight stages), the research involved 60 sixth-semester students. Instruments included expert validation, observation sheets, and learning outcome tests. Results showed that student and lecturer analyses confirmed the need for a systematic and creativity-oriented evaluation model. Expert validation achieved 86% (very appropriate); individual trials 89.6%, small group 86.8%, field trials 86.7%, practicality 85.3% (practical), and student response 86.7% (positive). Effectiveness testing showed a 17.5% improvement in learning outcomes with a medium effect size. The developed model is categorized as very valid, practical, and effective, making it feasible for wider implementation in engineering education to improve the quality and impact of learning evaluation practices.

Keywords: model development, learning evaluation, learning effectiveness

1 Introduction

Learning evaluation is a crucial component of the education system, serving as a means of assessing the achievement of learning objectives and providing feedback to both lecturers and students. In the context of higher education, particularly in the Mechanical Engineering Education study program, learning evaluation plays a strategic role in ensuring that students not only master cognitive aspects but also develop critical, analytical, and creative thinking skills in solving engineering problems [1].

Learning evaluation plays a strategic role in the education system, serving to ensure the achievement of instructional objectives, assess the quality of the learning process, and provide constructive feedback for the development of both students and lecturers [2]; [3]. In higher education, evaluation serves not only as an administrative tool for determining grades but also as a reflective instrument for improving the quality of learning and encouraging students to be more active, creative, and responsible in their learning process [4].

In the Mechanical Engineering Education study program, the Learning Outcome Evaluation course is a crucial pillar, equipping students with the skills to assess learning outcomes objectively, systematically, and with a focus on quality improvement. The ability to design evaluation instruments, analyze learning outcome data, and interpret evaluation results is essential for prospective educators and engineering practitioners to comprehensively measure student competency achievement [5].

Current evaluation practices tend to emphasize cognitive aspects and final outcomes (summative assessment), neglecting the holistic learning process and outcomes of students [6]. This has the potential to reduce learning effectiveness, as students receive less constructive feedback to improve their academic performance. Therefore, a more comprehensive and sustainable learning evaluation model is needed, based on learning analytics principles and an authentic approach that aligns with the characteristics of Learning Outcome Evaluation courses.

Learning evaluation practices in many universities still face several problems. First, lecturers tend to use traditional evaluation forms, such as multiple-choice tests or final exams, which emphasize lower-level cognitive abilities [7]. Second, there is a lack of diverse instruments capable of measuring students' psychomotor skills and professional attitudes, even though these skills are highly relevant to the characteristics of engineering learning [8]. Third, evaluation does not yet fully function as a learning tool (assessment for learning), but rather as an assessment of learning outcomes, thus limiting learning effectiveness [9].

Learning effectiveness itself is understood as the extent to which the learning process is able to produce positive changes in students' knowledge, skills, and attitudes in accordance with established instructional objectives [10]. In the context of mechanical engineering learning, effectiveness also includes active student involvement in laboratory practice, real-world problem-solving, and the application of theory to industrial practice [11]. Therefore, good evaluation must function not only to measure final results but also to monitor student progress throughout the learning process.

Various learning evaluation models have been introduced, such as the Kirkpatrick Model, which emphasizes four levels of learning outcomes [12], the comprehensive CIPP Model, which considers context, input, process, and product [13], and Authentic Assessment, which is more oriented toward real-world and contextual skills [14]. However, the implementation of these models still requires adaptation to suit the characteristics of mechanical engineering education, which demands a balanced integration of cognitive, affective, and psychomotor aspects.

In the context of Mechanical Engineering Education, learning effectiveness is closely related to the achievement of competencies that are not only theoretical, but also applicable. Appropriate evaluation will help students develop technical skills, systematic thinking, and reflective abilities in assessing their own learning processes and outcomes [15]. Thus, the development of an innovative and adaptive learning evaluation model is expected to increase the effectiveness of student learning and encourage the achievement of learning outcomes in accordance with the OBE (Outcome-Based Education)-based curriculum

However, the practice of learning evaluation in higher education still faces many obstacles. Several studies have shown that evaluations conducted by lecturers tend to focus on cognitive aspects through written tests, while students' affective and psychomotor aspects are under-recognized [16]; [17]. As a result, learning effectiveness is less than optimal because students do not obtain a complete picture of their competency development. Learning effectiveness is measured not only by grades but also by students' ability to integrate knowledge with practical skills in mechanical engineering [18].

To address these challenges, the development of a more comprehensive, authentic, and adaptive learning evaluation model is necessary. This evaluation model must be able to integrate various approaches, such as formative, summative, portfolio, project-based assessment, and learning analytics, to produce more accurate and useful information for both lecturers and students [19]. With the right evaluation model, lecturers can design more effective learning strategies, while students receive feedback that supports the continuous improvement of their learning quality.

In line with the Outcome-Based Education (OBE) paradigm, learning evaluation in Mechanical Engineering Education must ensure that students not only understand the concept of learning outcome evaluation but also apply this knowledge in real-world contexts, both as prospective educators and as engineering practitioners [20]. Therefore, research on the development of learning evaluation models is urgently needed to improve the effectiveness of student learning in the Learning Outcome Evaluation course.

Based on this background, research on the development of learning evaluation models in the Learning Outcome Evaluation course in Mechanical Engineering Education is crucial. The goal is to produce an evaluation model that can improve the effectiveness of student learning, encourage active involvement in the learning process, and support the achievement of graduate competencies that meet the demands of education and industry.

2 Method

This research uses a research and development (R&D) method. This approach was chosen because the primary objective of the research is to produce a valid, practical, and effective learning evaluation model to improve the effectiveness of student learning in the Learning Outcome Evaluation course in the Mechanical Engineering Education Study Program.

The development model used refers to the steps of Borg & Gall [20], simplified as needed, with the following stages: Needs analysis, Model design, Expert validation, Revision, Limited trial, Extensive trial, Practicality and response testing, Effectiveness testing.

Research object: Learning evaluation model for the Learning Outcomes Evaluation course. Research subjects: (1) Learning evaluation subject matter experts (educational assessment experts); (2) Learning design experts (educational technology lecturers); (3) Mechanical Engineering Education lecturers who teach the Learning Outcomes Evaluation course; and (4) Sixth-semester Mechanical Engineering Education students (trial class). Students: 60 sixth-semester Mechanical Engineering Education students. 30 students in the experimental group (using the developed evaluation model); 30 students in the control group (using conventional evaluation).

Data Collection Techniques: Data were obtained through: (1) Observation: implementation of learning and evaluation; (2) Interviews: lecturers and students regarding the weaknesses of the current evaluation model; (3) Questionnaires: expert validation, lecturer and student responses; (4) Learning Outcome Test: student pretest–posttest; (5) Documentation: Lesson Plans (RPS), previous evaluation instruments, student learning outcomes; (6) Learning Analytics: student activity data from the LMS/Google Classroom digital platform (material access, participation, online attendance).

Expert Validation: Descriptive analysis (average score & eligibility category). Model Effectiveness: N-Gain test to determine the increase in student creativity before and after model implementation.

$$g = \frac{\text{Posttest Score} - \text{Pretest Score}}{\text{Maximum Score} - \text{Pretest Score}} \quad (1)$$

N-Gain Categories: High (≥ 0.70), Medium (0.30–0.69), Low (< 0.30).

The results of the feasibility and practicality test can be calculated using the following formula:

Note:

$$\text{Percentage} = \frac{\sum \text{score per item}}{\sum \text{maximum score}} \times 100\% \quad (2)$$

The expert validation results are then adjusted according to the criteria in the following table:

Table 1. Conversion of Product Feasibility and Practicality Levels

Achievement Level (%)	Category	Description
81 – 100	Very Feasible/Practical	No need for revision
61 – 80	Feasible/Practical	No need for revision
41 – 60	Fairly Feasible/Practical	Revised
21 – 40	Less Feasible/Practical	Revised
0 – 20	Not Feasible/Practical	Revised

The learning test results are assessed based on scoring guidelines. The analysis is carried out in the following stages: (1) assigning a score to each student's answer based on the assessment rubric that has been created. (2) adding up the scores obtained by the students. (3) calculating the score obtained by each student. (4) categorizing the student learning outcome test according to the school's criteria. (5) tabulating the student test results. (6) calculating the percentage of student test completion using the formula:

$$\text{Percentage (X)} = \frac{\text{number of students who completed}}{\text{number of students (n)}} \times 100\% \quad (3)$$

(7) Converting the learning outcome test data to the learning outcome effectiveness guideline Table 2.

Table 2. Effectiveness Table

Percentage of Completion (%)	Effectiveness
$X > 80$	Very Good
$60 < X \leq 80$	Good
$40 < X \leq 60$	Sufficient
$20 < X \leq 40$	Poor
$X \leq 20$	Very Poor

Statistical Analysis: (1) Instrument reliability test (Cronbach's $\alpha > 0.7 \rightarrow$ reliable); (2) Normality and homogeneity test before the t-test; (3) N-Gain to see score improvement; (4) Effect size to assess the strength of the influence. Data Analysis Techniques: (1) Descriptive statistics (validation and feasibility); (2) N-Gain test; (3) t-test (independent and paired); (4) Effect size calculation.

3 Results and Discussion

3.1 Research Results

Validation was conducted by three experts: a learning materials expert, a learning media expert, and a learning design expert. Each expert assessed the content, construct, and design aspects using a Likert scale (1–4).

Table 3. Expert Validation Results

Validation Aspect	Material Expert	Media Expert	Design Expert	Average	Aiken's V
Content suitability	3.6	3.7	3.5	3.6	0.90
Construct feasibility	3.5	3.6	3.4	3.5	0.88
Model design clarity	3.7	3.6	3.6	3.63	0.91
Average Total	3.6	3.63	3.5	3.58	0.90

Results: Aiken's V = 0.90 (≥ 0.70), indicating that the model is highly valid..

Effectiveness Test

Effectiveness was tested using a pretest-posttest control group design on 60 students (30 experimental, 30 control).

Table 4. Model Effectiveness Test Results

Group	N	Mean Pretest	Mean Posttest	N-Gain	Category	t-test (p)	Effect Size (d)
Experimental	30	62.3	83.7	0.58	Medium	0.000*	0.85 (high)
Control	30	61.9	72.5	0.35	Low		

Note: * $p < 0.05$ is significant. Results: The experimental group improved by 21.4 points (62.3 \rightarrow 83.7). The control group only improved by 10.6 points (61.9 \rightarrow 72.5). N-Gain for the experimental group = 0.58 (medium), and for the control group = 0.35 (low). The t-test showed

a significant difference ($p = 0.000$). Effect size = 0.85 (large) → the model is very effective in improving learning outcomes.

Table 5. Summary of Research Results

Research Stage	Main Results	Conclusions
Needs analysis	70–87% weaknesses in the old evaluation	Need for a new model
Expert validation (3 experts)	Aiken's $V = 0.90$	Model is very valid
Individual trial	82.5%	Good
Small group trial	85%	Very good
Main field trial	87.5%	Very good
Practicality test	84.3%	Very practical
Student response	83.3%	Positive
Effectiveness test	N-Gain 0.58, ES = 0.85	Effective

Conclusion: The developed learning evaluation model is very valid ($V = 0.90$), practical (84.3%), received a positive response (83.3%), and has been proven effective (N-Gain = 0.58; Effect size = 0.85) in improving the learning outcomes of Mechanical Engineering Education students.

3.2 Discussion

The results of the needs analysis indicate that learning evaluation in the Evaluation of Learning Outcomes course is still dominated by written tests (86.7%), psychomotor aspects are under-evaluated (73.3%), feedback is limited (66.7%), and there is no use of learning analytics-based evaluation technology (70%). These findings align with Arikunto's [21] view that educational evaluation practices in Indonesia still tend to emphasize final outcomes in the form of cognitive tests. This creates a gap, as engineering learning should also emphasize the assessment of practice and skills.

Feisel & Rosa [22] emphasized that in engineering education, assessment must address pre-lab (planning), in-lab (practice), and post-lab (analysis & reflection) aspects. Therefore, the need for a comprehensive, adaptive, and technology-based evaluation model is highly relevant.

Validation results by three experts showed an Aiken's V value of 0.90, indicating the model is highly valid. This demonstrates that the content, construct, and design aspects of the model align with the principles of learning evaluation. According to Messick [23], validity is the primary foundation of assessment instruments, as it relates to the validity of the interpretation of evaluation results.

Nitko & Brookhart [24] emphasized that a valid evaluation instrument will ensure that assessment results truly reflect student abilities. This finding is also consistent with research by Putra & Kustiono [25], which demonstrated that expert validation plays a crucial role in testing the suitability of performance-based evaluation instruments for vocational education.

Expert validation results (Aiken's $V = 0.91$) confirmed that the developed evaluation model was suitable for use because it adhered to the principles of content, construct, and design validity [26]. This aligns with Nitko & Brookhart's [27] opinion, which states that valid evaluation instruments are a prerequisite for ensuring learning quality.

The results of individual (82.5%), small group (85%), and main field (87.5%) trials indicate that the developed evaluation model was practical for students to use. The increase in scores from individual to main field trials indicates an improvement in the model's quality after revisions based on student feedback.

These findings support Borg & Gall's [28] theory that every educational development product must undergo multi-stage trials (individual, small group, and field) to produce a final product that is valid, practical, and meets user needs.

The results of the practicality test indicate that the model is easy to use by both lecturers and students, with an average score of 84.3% (very practical). These findings are consistent with research by Putra & Kustiono [29], which stated that performance rubric-based evaluation instruments are easier to understand and use in vocational learning.

Student responses were also very positive (83.3%), particularly regarding motivation and satisfaction. This reinforces Earl's theory of assessment for learning [30], which states that evaluations can increase learning motivation if designed authentically and relevantly.

According to Sukardi [31], the practicality of evaluation instruments is important so that lecturers can use them without reducing learning time and students can understand them without confusion. These findings also align with research by Brookhart [32], which emphasized that clear and simple assessment rubrics will increase the practicality and objectivity of assessments.

Student responses to the evaluation model were very positive (83.3%), particularly regarding motivation (84.5%) and satisfaction (85.9%). This indicates that students feel more motivated and satisfied with evaluations that assess various aspects of competency, not just test scores.

Earl [33] explains that good assessments function as assessment for learning and assessment as learning, encouraging students to learn better through feedback. Research by Black & Wiliam [34] also supports the belief that integrated formative evaluation can increase student motivation and engagement.

The effectiveness test results showed that the experimental group experienced a higher increase in learning outcomes (N-Gain = 0.58, moderate category) compared to the control group (N-Gain = 0.35, low category). This difference was significant ($p = 0.000$), with an effect size of 0.85 (large category). This indicates that the evaluation model effectively improves student learning outcomes. This aligns with research by Black & Wiliam [34], which demonstrated that formative evaluation significantly improves learning outcomes.

These findings are consistent with Hattie's theory [35], which, through meta-analysis, stated that feedback and formative assessment have a significant influence on learning outcomes ($d = 0.73$). Research by Prince & Felder [36] also demonstrated that project-based assessment in engineering education can improve critical thinking and problem-solving skills.

The integration of learning analytics into the model also strengthens its effectiveness. Siemens & Long [37] emphasized that learning analytics can provide real-time data to monitor student progress, enabling more targeted learning interventions. This explains why students in the experimental group were more engaged and achieved higher levels of achievement than those in the control group.

The integration of authentic assessment and learning analytics in this model has also been shown to strengthen student engagement, as emphasized by Wiggins [38] and Siemens & Long [39]. Thus, this model supports the Outcome-Based Education (OBE) principle, which emphasizes alignment between CPL, process, and evaluation [40].

The e-learning-based collaborative development model has been proven to be feasible, practical, and effective in improving students' work skills learning outcomes, with the evaluation score for the collaboration aspect reaching the 'Very Good' category ($\geq 90\%$) [41].

The effectiveness of a learning model is influenced by student characteristics, so the choice of model needs to be adjusted to individual differences, especially the level of creative thinking ability [42].

4 Conclusion

Based on the research results and discussion regarding the Development of a Learning Evaluation Model to Improve Student Learning Effectiveness in the Learning Outcome Evaluation Course in Mechanical Engineering Education, the following conclusions can be drawn:

The needs analysis shows that the learning evaluation used by lecturers still focuses primarily on cognitive aspects, while students' affective and psychomotor aspects have not been optimally accommodated. This indicates the need to develop a more comprehensive learning evaluation model.

Expert validation (material experts, learning media experts, and learning design experts) showed an average result of Aiken's $V \geq 0.90$, which means that the developed learning evaluation model is very valid and worthy of implementation with minor improvements.

Individual, small group, and main field trials demonstrated increased student acceptance. In the main field trial, the model achieved a practicality score of 84.3% (very practical), indicating effective use by both lecturers and students.

Student response to the use of the learning evaluation model was very positive, with an average score of 83.3%, particularly in the areas of motivation, learning satisfaction, and active student engagement in the learning process.

Effectiveness testing using a pretest-posttest control group design demonstrated a significant difference between the experimental and control groups ($p < 0.05$). The experimental group experienced improved learning outcomes with an N-Gain of 0.58 (moderate category) and an effect size of 0.85 (large category). This demonstrates that the developed evaluation model is effective in enhancing student learning effectiveness in the Learning Outcome Evaluation course in Mechanical Engineering Education.

Acknowledgments.

The author would like to express his deepest gratitude to Medan State University for funding this research through the 2025/2026 Non-Tax State Revenue (PNBP) funds, in accordance with the Decree of the Rector of Medan State University Number: 0271/UN33/KPT/2025. This funding support plays a crucial role in the smooth running and completion of this applied product research.

References

- [1] Nitko, A. J., & Brookhart, S. M. *Educational Assessment of Students*. Boston: Pearson. (2014).
- [2] Tyler, R. W. *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press. (2013).
- [3] Nitko, A. J., & Brookhart, S. M. *Educational Assessment of Students*. Boston: Pearson. (2014).
- [4] Wiggins, G. *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco: Jossey-Bass. (1998).
- [5] Mursid, R. *Pengembangan Model Pembelajaran dalam Pendidikan Teknik dan Kejuruan*. Bandung: Alfabeta. (2016).
- [6] Sukardi. *Evaluasi Pendidikan: Prinsip dan Operasionalnya*. Jakarta: Bumi Aksara. (2015).
- [7] Arikunto, S. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara. (2019).
- [8] Brown, S., & Glasner, A. *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*. Maidenhead: Open University Press. (2003).
- [9] Black, P., & Wiliam, D. Developing the Theory of Formative Assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. (2009).
- [10] Slavin, R. E. *Educational Psychology: Theory and Practice*. Boston: Pearson. (2018).
- [11] Joyce, B., Weil, M., & Calhoun, E. *Models of Teaching*. Boston: Pearson. (2015).
- [12] Kirkpatrick, D. L. Techniques for Evaluating Training Programs. *Journal of the American Society of Training Directors*. (1959).
- [13] Stufflebeam, D. L. The CIPP Model for Evaluation. In Kellaghan, T., & Stufflebeam, D. L. (Eds.), *International Handbook of Educational Evaluation* (pp. 31–62). Springer. (2003).
- [14] Wiggins, G. *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco: Jossey-Bass. (1998).
- [15] Brown, S., & Glasner, A. *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*. Maidenhead: Open University Press. (2003).
- [16] Arikunto, S. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara. (2019).
- [17] Brown, S., & Glasner, A. *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*. Maidenhead: Open University Press. (2003).
- [18] Joyce, B., Weil, M., & Calhoun, E. *Models of Teaching*. Boston: Pearson. (2015).
- [19] Siemens, G., & Long, P. Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46(5), 30–40. (2011).
- [20] Borg, W. R., & Gall, M. D. *Educational Research: An Introduction*. New York: Longman. (2003).
- [21] Arikunto, S. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara. (2019).
- [22] Feisel, L. D., & Rosa, A. J. The role of the laboratory in undergraduate engineering education. *Journal of Engineering Education*, 94(1), 121–130. (2005).
- [23] Messick, S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry. *American Psychologist*, 50(9), 741–749. (1995).
- [24] Nitko, A. J., & Brookhart, S. M. *Educational Assessment of Students*. Boston: Pearson. (2014).
- [25] Putra, A., & Kustiono. Authentic assessment for vocational education: Developing performance-based evaluation instruments. *Jurnal Pendidikan Vokasi*, 10(2), 150–162. (2020).
- [26] Messick, S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry. *American Psychologist*, 50(9), 741–749. (1995).
- [27] Nitko, A. J., & Brookhart, S. M. *Educational Assessment of Students*. Boston: Pearson. (2014).
- [28] Borg, W. R., & Gall, M. D. *Educational Research: An Introduction*. New York: Longman. (2003).
- [29] Putra, A., & Kustiono. Authentic assessment for vocational education: Developing performance-based evaluation instruments. *Jurnal Pendidikan Vokasi*, 10(2), 150–162. (2020).
- [30] Earl, L. *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*. Corwin. (2003).
- [31] Sukardi. *Evaluasi Pendidikan: Prinsip dan Operasionalnya*. Jakarta: Bumi Aksara. (2015).
- [32] Brookhart, S. M. *How to Create and Use Rubrics for Formative Assessment and Grading*. ASCD. (2013).
- [33] Earl, L. *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*. Corwin. (2003).

- [34] Black, P., & Wiliam, D. Developing the Theory of Formative Assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. (2009).
- [35] Hattie, J. *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge. (2009).
- [36] Prince, M. J., & Felder, R. M. Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education*, 95(2), 123–138.
- [37] Siemens, G., & Long, P. Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46(5), 30–40. (2011).
- [38] Wiggins, G. *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco: Jossey-Bass. (1998).
- [39] Siemens, G., & Long, P. Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46(5), 30–40. (2011).
- [40] Biggs, J., & Tang, C. *Teaching for Quality Learning at University* (4th ed.). Open University Press. (2011).
- [41] Mursid, R., Muslim, & Fariyah. (2023). Collaboration-based development model e-learning on course learning achievements working skills. *International Journal of Instruction*, 16(2), 307–328. <https://doi.org/10.29333/iji.2023.16218a>
- [42] Mursid, R., Saragih, A. H., & Hartono, R. (2022). The effect of the blended project-based learning model and creative thinking ability on engineering students' learning outcomes. *International Journal of Education in Mathematics, Science, and Technology (IJEMST)*, 10(1), 218–235. <https://doi.org/10.46328/ijemst.2244>