DOI:https://doi.org/10.52131/jer.2025.v7i1.3094

iRASD Journal of Educational Research

Volume 7, Number 1, 2026, Pages 01 - 16

Journal Homepage:
https://journals.internationalrasd.org/index.php/jer

iRASD JOURNAL OF EDUCATIONAL RESEARCH

INTERNATIONAL RESEARCH ASSOCIATION FOR SUSTAINABLE DEVELOPMENT

The Impact of Alternative Assessment on Non-Affective Learning Outcomes: A Meta-Analysis of Experimental and Quasi-Experimental Studies

Haneen Taqatqa [1], Abdalkarim Ayyoub [2]

[1] Physics Teacher/Ph.D. Scholar, Al-Quds Open University/An-Najah National University, Palestine.
Email: haneenmtaqatqa@gmail.com

[2] Professor, Faculty of Education and Teacher Preparation, An-Najah National University, Palestine.
Email: ayyoub@najah.edu

ARTICLE INFO

Article History:

Received:	February 10, 2025
Revised:	April 24, 2026
Accepted:	April 25, 2026
Available Online:	April 26, 2026

Keywords:
Alternative Assessment
Learning Outcomes
Self-Assessment
Peer Assessment
Technology-Based Assessment
Meta-Analysis

Funding:
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

OPEN ACCESS

ABSTRACT

With the increasing interest of researchers in alternative assessment methods, their widespread adoption, and the abundance of related research, this study aims to compile and analyze the findings of these studies comprehensively to arrive at a holistic conclusion and provide a comprehensive overview of their results. Specifically, this study seeks to reveal the overall impact of using alternative assessment on non-affective learning outcomes. It also investigates the varying impact sizes of different types of alternative assessment (self-assessment, peer assessment, and technology-based assessment) on these non-affective learning outcomes. Therefore, the researcher relied on meta-analysis, collecting numerous published research papers and studies according to a set of criteria obtained from a literature review, and using the PRISMA framework. After identifying the research sample, the risk of bias was assessed using Cochrane tools, excluding studies classified as high risk. A precise analysis card, developed by the researcher after confirming its validity and reliability, was then used. This card was initially used for qualitative analysis of the research and then converted to quantitative values to facilitate the use of the data in the meta-analysis. The results showed that the overall effect size of using alternative assessment on non-affective learning outcomes was moderate (0.44). The results also showed no statistically significant differences between effect sizes attributable to the assessment method used; that is, the results did not reveal differences in effect sizes between self-assessment, peer assessment, and technology-supported assessment on non-affective learning outcomes. Therefore, there remains an unexplained variability between the studies, requiring further research to explore the moderating factors.

Corresponding Author's Email: haneenmtaqatqa@gmail.com

1. Introduction

Alternative assessment is an assessment approach that immerses learners in rich experiences in educational contexts, where students demonstrate what they have learned through purposeful and realistic activities that go beyond memorization and recall. In contrast, traditional assessment relies heavily on tests that often prioritize memorization. Alternative assessment stems from the need to move beyond this narrow traditional measurement to a broader form that values the learner's competence and ability to transfer and apply knowledge in authentic and complex contexts (Meylani, 2024). It also accommodates diverse cognitive and skill patterns, opening avenues in line with the development of metacognitive skills and the self-regulation of students' abilities and potential, while acknowledging their individuality and understanding their differences. In contrast, traditional methods marginalize individual learning differences among students (Panadero, Jonsson, & Botella, 2017).

Thus, alternative assessment is not only a way to measure learning more comprehensively, but also represents a learning experience in itself that promotes independence, critical inquiry and deep understanding in the student. Alternative assessment focuses on the authenticity of the results and the learner's role in shaping the assessment process itself. It does not only evaluate the final result but also evaluates the cognitive path and learning strategies, encouraging students to demonstrate their understanding through purposeful performance in active interactive environments (Vallis, 2025). The increasing integration of alternative assessment methods in educational environments extending from the primary stage to higher education has coincided with the interest in major shifts in educational thought, represented by moving away from teaching patterns and models based on knowledge transfer towards constructivist models, which emphasize the active role of students in building their learning outcomes. This shift supports not only the acquisition of content, but also the development of critical thinking abilities, self-regulation skills, and assessment competencies that go beyond academic contexts (Nisbet & Shaw, 2024). The contemporary educational landscape has unprecedented requirements and challenges that have contributed to highlighting the importance of alternative assessment methods, as educational institutions serve student groups that come from diverse and completely different cultural, linguistic, social and economic backgrounds, as well as their educational needs in the same educational situation. These gaps and differences are increasing, which widens the gaps between their levels. Therefore, this diversity requires assessment frameworks that are able to accommodate them while maintaining the accuracy and credibility of the assessment process. Alternative assessment methods, in their various forms, are promising methods in this regard, as they provide flexible assessment methods that can be adapted in diverse contexts (Sukumal, 2024).

In this renewed educational context, the impact of alternative assessment methods on learning outcomes has become a focal point of significant research interest. Intuitive links have been proposed between student effectiveness in assessment and improved learning outcomes (Quinlan, Sellei, & Fiorucci, 2025), based on theoretical frameworks indicating that student interaction with assessment criteria, peers, or even hypothetical samples they evaluate allows them to generate or receive feedback from the learning environment that can help them deepen their understanding and improve their performance (Luan & Narayanan, 2024). Because alternative assessment methods are among the most important educational practices that educational reality has prompted a decision regarding their importance and impact on students' learning outcomes, they have also received great attention stemming from modern theoretical foundations. Constructivist theory portrays the educational process as knowledge that is built internally through active engagement and interaction in meaningful educational contexts, helping the student to adapt their existing cognitive frameworks to the construction of new knowledge that they acquire (Ahmad, Sultana, & Jamil, 2020). From this perspective, alternative assessment methods can be understood as mechanisms that facilitate the construction process by encouraging students to take responsibility for their learning. According to this vision, alternative assessment is part of the learning process. Through self-assessment and peer assessment, for example, the student is given an opportunity that enhances their monitoring of the development of the learning process, which helps them to establish knowledge and consolidate understanding (Wakeman et al., 2021).

In the same context, Vygotsky focused on dynamic assessment, which focuses on social interaction as a means to drive the cognitive and performance development of the student by focusing on the participatory and dialogical nature of the knowledge-building process, especially in the area of imminent growth (Raslan, 2024), Assessment through peers or engaging with them in specific tasks, for example, enhances communication between them and increases the exchange of experiences and intensifies attention. It can also be a supportive and motivating tool that enables the student to surpass his current level to higher levels (Daneshfar & Moharami, 2018). Self-regulated learning theories such as Bandura's theory and Zimmerman's theory also provide another theoretical framework for understanding the possible effects of using alternative assessment methods. They portray learning as a process in which students diagnose and direct their perception, motivation, and behavior toward achieving goals, after going through periodic assessment cycles, relying on self-observation, self-reflection, evaluation, and self-criticism in a way that enhances their performance Flavell (2024); Yan (2020) also emphasized the importance of metacognitive awareness in improving academic achievement, as it is one of the foundations of alternative assessment methods that require the student to actively engage in performing educational tasks that are not intended to detect errors, but rather are an educational means that stimulates their learning and achievement, thus providing a strong justification for the importance of relying on alternative assessment in education.

With the shift towards using technological tools in the assessment process as one of the modern alternative assessment methods, teachers were able to overlook the negative aspects shown by alternative assessment methods based on students, such as bias and competition among students, or the lack of clarity in assessment criteria and other problems that affect the measurement process (Corbin et al., 2025). In contrast, the absence of the human element remains a real problem, as ignoring students' personalities and psychological and social factors has a significant negative impact when relying on technology and its tools, such as artificial intelligence, in the assessment process. This actually creates an automated learning environment, in addition to other concerns related to data privacy (Seth & Sankarapu, 2025). Despite this, the empirical reality remained complex and sometimes contradictory. While many studies indicated positive correlations between alternative assessment methods and student academic performance and cognitive development, others revealed results that raised doubts about their effectiveness, indicating that the effectiveness of these methods is influenced by various factors and factors, most notably research methodologies, student characteristics, subject areas, institutional contexts, and others (Monib, Karimi, & Nijat, 2020). In light of this, this study came to investigate, through meta-analysis, the impact of alternative assessment on learning outcomes, focusing specifically on research with experimental and quasi-experimental designs, which provide the strongest evidence of causal relationships, especially in light of the increasing need to make evidence-based educational decisions to adopt reliable educational practices in the field of assessment (Leppink, 2019). The reliance on the use of meta-analysis as a research methodology has increased, which relies on the results of available research to reach a deeper and broader understanding through the data available in published research (Maynard, 2024), In hence the research questions:

- What is the overall effect size of using alternative assessment methods on non-affective learning outcomes?
- Do the effect sizes differ depending on the type of alternative assessment used in the studies included?

This study aims to achieve a set of interconnected and sequential objectives through in-depth analysis. It begins by understanding the overall impact of alternative assessment methods on non-affective learning outcomes, encompassing cognitive and skill-based aspects. The study then gradually expands to include more detailed aspects. In general, the study aims to reveal the extent of the impact of using alternative assessment methods (self-assessment, peer assessment, and digital technology-based assessment) on non-affective learning outcomes. This is achieved through a comprehensive quantitative analysis that enhances understanding of the reliability of these methods and any differences between them. This provides an opportunity to compare these methods and identify the most influential approach on learning outcomes. Therefore, the study seeks to explore whether the type of assessment is a moderating variable that affects effect sizes. Ultimately, this contributes to understanding the conditions and contexts that influence the effectiveness of alternative assessment. So this study derives its importance from two dimensions:

First: Theoretical Significance

This study makes a significant contribution to educational theory, adding value to the growing body of knowledge about alternative assessment as a well-established educational approach supported by several learning theories, most notably constructivism. It also reinforces the theoretical foundations that emphasize learning as a social process, thus providing a robust theoretical framework for understanding how alternative assessment practices align with contemporary learning theories that prioritize constructing purposeful knowledge over passively absorbing information. Through a rigorous and comprehensive meta-analysis of experimental and quasi-experimental studies spanning decades, this study addresses a critical gap in the literature by integrating experimental evidence to draw a comprehensive conclusion from a collection of fragmented and sometimes inconsistent findings. This integration is theoretically significant, as it transcends the limitations of individual research, providing a more robust and generalizable understanding of the impact of alternative assessment on learning outcomes.

Second: Practical Significance

The practical implications of this study extend to numerous stakeholders and educational contexts. It provides teachers with concrete, evidence-based insights, derived from a rigorous meta-analysis, enabling them to make informed decisions about the assessment methods that most effectively impact non-affective learning outcomes. This evidence-based approach supports professional decision-making and helps teachers justify their choices to administrators, parents, and students. Furthermore, this study offers educational policymakers at the local, regional, and global levels crucial evidence for developing assessment policies based on empirical research, rather than unsubstantiated claims. The comprehensive nature of the meta-analysis empowers policymakers to confidently implement alternative assessment initiatives, allocate resources for professional development, and establish assessment frameworks that support meaningful learning outcomes. Curriculum developers and program designers can also benefit from this study by identifying more targeted, relevant, and impactful assessment methods and practices that positively influence learning outcomes.

2. Research Literature

Alternative assessment is a relatively recent educational trend that dates back to the 1970s, emerging as a reaction to the limitations of traditional tests that focus on memorization and recall (Jacobs & Renandya, 2019), by seeking to provide a comprehensive picture of students' abilities through assessing their skills in real educational situations. This concept has spread as a necessity within attempts to reform traditional assessment systems (Demir, 2021). The features of alternative assessment were defined during the 1980s, as the real beginning of this concept was in the United States of America when educators and researchers began to question the effectiveness of traditional tests and assessments in measuring students' real abilities, as calls increased for the development of assessment methods that promote deep learning and support logical thinking (Black & Wiliam, 2018). At the time, research bodies contributed to establishing this concept and linking it to school reform, such as the National Center for Student Assessment, Standards and Testing Research (CRESST), which was established in the United States in partnership with a number of universities such as Cambridge, California, Stanford and others in the mid-1980s, with funding from the US Department of Education, with the aim of setting educational standards that improve assessment methods, and supporting schools and teachers to formulate fair assessment policies (Bechu et al., 2024).

The initial phase of alternative assessment development focused on developing assessment methods that emphasized processes rather than solely final results. There were also growing calls for more comprehensive and in-depth assessment methods that supported student participation and interaction (Nieminen, Haataja, & Cobb, 2025), leading to the first serious attempts to implement assessment methods different from traditional tests. This development was driven by several educational, social, and cultural factors, most notably the shift in the concept of education from rote learning to the active construction of knowledge, and the increasing attention given to individual student differences (Black & Wiliam, 2018). The 1990s witnessed a rapid growth in the application of the alternative assessment concept. Herman (1992) emerged as one of the first practical guides aimed at helping teachers understand the integration of alternative assessment into classroom practices. The book highlighted the importance of reconsidering assessment and the necessity of linking assessment to teaching. It also identified a set of assessment criteria to guide the design and implementation of alternative assessment tools. As a result, alternative assessment became more structured and systematic, and educators began to establish theoretical and practical foundations for its effective application. Numerous alternative assessment methods and approaches emerged and gained widespread acceptance, contributing to the consolidation of the concept of alternative assessment as a supportive assessment method for active and purposeful learning in contemporary educational practices (Meylani, 2024).

Integrating alternative assessment has significantly contributed to enhancing critical thinking and developing students' ability to apply knowledge in new, real-world situations. It also provides valuable feedback that helps teachers improve their instruction (Shavelson et al., 2019). With the dawn of the new millennium, alternative assessment underwent a remarkable transformation due to rapid technological advancements. Teachers began using advanced digital tools to implement more interactive and diverse assessment methods. This development enabled the application of new assessment methods to measure more skills more accurately and easily. This evolution was influenced by globalization and the evolving demands of the labor market resulting from the need for new skills (Maki, 2023). Research has increasingly focused on its effectiveness and role in verifying its impact on learning outcomes. Educators are now shifting their focus to various alternative assessment methods to achieve a more balanced, comprehensive, and equitable system (Demir, 2021). This phase has witnessed a greater understanding of alternative assessment and its applications, to the point where it is now described as a complete alternative that can replace traditional assessment, not merely a supplement to it (Swiecki et al., 2022).

These successive transformations and the introduction of technology have led to the use of artificial intelligence in the assessment process, as it has increased the accuracy and speed of analyzing student performance in an exceptional way, and provided personalized assessments that suit each student, with accurate monitoring that enhances the effectiveness of the educational process and stimulates the continuous development of all students (Shapiro et al., 2025). It can also contribute to reducing bias and increasing objectivity, which contributes to creating a more equitable and inclusive environment. However, the integration of artificial intelligence still needs to be done with caution and deliberation to ensure the quality of the assessment without neglecting the human dimension in the educational process, as it is essentially a social process (González-Calatayud, Prendes-Espinosa, & Roig-Vila, 2021). Peer assessment is a prominent form of alternative assessment, based on students' active participation in evaluating their peers' performance according to predefined criteria. The importance of this type of assessment lies in its ability to develop students' critical thinking skills through analyzing and evaluating their peers' work (Fleckney, Thompson, & Vaz-Serra, 2025). This enhances their understanding of required academic standards and develops their ability to conduct accurate and objective evaluations. This, in turn, contributes to strengthening students' sense of responsibility and encourages collaborative learning through the exchange of experiences and knowledge, creating an interactive learning environment that fosters continuous progress and development (Quinlan, Sellei, & Fiorucci, 2025).

As a result of this positive interaction among students, the concept of assessment evolves to capture their attention and increase their focus on the subjective dimension, which centers on the student's ability to evaluate their own performance. Self-assessment is deeply rooted in self-awareness, enhancing students' capacity for self-reflection and identifying their strengths and weaknesses (Sinaga et al., 2024). This necessitates providing appropriate and clearly defined assessment tools that guide students during the required tasks and enable them to objectively evaluate their performance (Nieminen & Boud, 2025). Self-assessment also demonstrates learners' ability to track their academic progress and level of achievement, making them active participants in the evaluation process. This fosters student autonomy and enhances their intrinsic motivation to learn, as they become more aware of their learning process and better equipped to identify appropriate learning strategies to improve their performance. Furthermore, self-assessment contributes to building self-confidence and developing personal planning and organizational skills, preparing students to face challenges (Andrade, 2019). This highlights the importance of students documenting their achievements, as it provides them with an opportunity to reflect on these accomplishments, monitor their development over successive periods, and compare their past performance with their current one in a cumulative manner. This method is considered a comprehensive alternative assessment that compiles student work to provide a complete and integrated picture of their learning level and achieved academic growth. It allows students to showcase their abilities and skills in diverse and innovative ways, and it helps teachers track their students' progress and provide feedback based on the evidence and data retained by the students (Naka, 2025).

In conjunction with rapid technological development and its penetration into various aspects of life, the reliance on digital technology applications in assessment has emerged as one of the modern methods. Research continues to focus on examining the impact of its use on all elements and outcomes of the educational process. This type of assessment is considered a true revolution in the field of education, based on advanced applications and machine learning techniques for analyzing and accurately and comprehensively evaluating student performance in a way that is characterized by objectivity and consistency (Bulut & Beiting-Parrish, 2024). In addition, artificial intelligence has recently emerged, and published research has shown its great excellence in the educational field, and its use in assessment as well, as it has the ability to deal with a large amount of data and process and analyze it at high speed, which gives an opportunity to provide immediate feedback that helps students know and understand their level of performance and identify areas that require development and improvement (Fang, Roscoe, & McNamara, 2023). The assessment process with artificial intelligence also ensures obtaining objective results that are free from bias. It is also distinguished by its ability to personalize education and assessment, as it can analyze different learning styles and offer diverse assessments that are consistent with students' abilities and characteristics, which contributes to saving time and effort.

3. Research Methodology

This study adopted a mixed-methods meta-analysis approach. It began with a qualitative analysis of the studies included in the sample, followed by a direct qualitative content analysis with quantification. Initially, the analysis was deductive, as this step was crucial to ensuring the validity of the transition to quantitative analysis. This deductive approach helped avoid combining results that were conceptually or methodologically incompatible (Çoğaltay & Karadağ, 2015). The primary objective was not to identify common categories or themes across the research or to develop quantifiable hypotheses, but rather to obtain more detail to ensure accurate data classification and facilitate subsequent interpretation of the results. However, the analysis revealed some categories that were added to the analysis scorecard, thus shifting the analysis from deductive to inductive. In the second stage, the results obtained were transformed into quantitative data by calculating effect sizes and coding the qualitative results numerically, in preparation for conducting the meta-analysis. By combining the results of several related non-meta-analytical studies, a more comprehensive and generalizable understanding can be provided (Borenstein et al., 2021). This methodological sequence is based on literature that emphasizes that initial qualitative analysis enhances the validity of meta-analysis, improves the interpretation of its results, and reduces the variance resulting from contextual differences between studies. Conducting an initial qualitative analysis of the research included in the sample is an essential step to ensure conceptual and methodological homogeneity before moving to quantitative analysis (Levitt, 2024). Accordingly, the researcher believes that combining qualitative and quantitative analysis was not a complementary procedure, but rather a justified and necessary methodological choice.

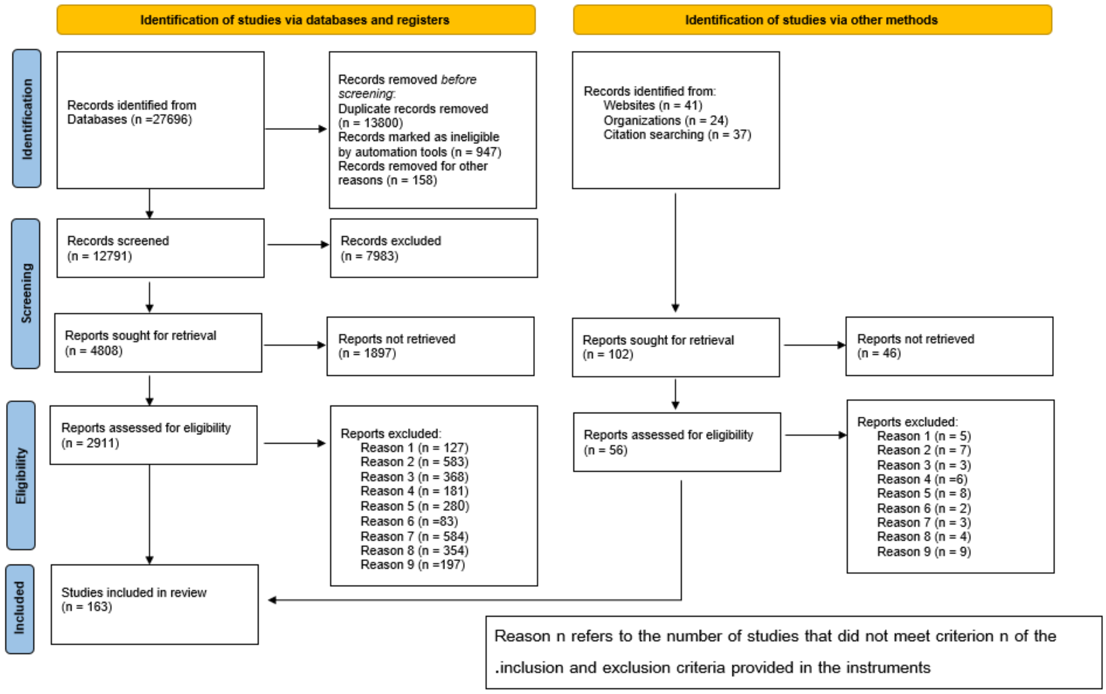
3.1. Study Population

The study population was defined as research published in peer-reviewed journals, master's and doctoral theses published in university repositories, and conference papers available on Google Scholar. This population consisted of published research examining the impact of alternative assessment on learning outcomes, obtained by searching for the keywords specified in the appendices. The total number of research papers was 27,696. The results of all searches conducted in RIS format were transferred to the Rayyan.ai platform (https://www.rayyan.ai/) to be compiled into a single database. This platform is designed to accelerate and organize the process of conducting systematic literature reviews. It provides an advanced system for sorting studies based on machine learning algorithms, which contributes to reducing sorting time by learning the researcher's decision patterns and providing recommendations based on them. It also provides automatic duplication removal by recognizing and excluding duplicate research without manual intervention (Wu & Chang, 2025). A study by Ng et al. (2024) showed that the sensitivity of this platform to detect relevant research was 96%-100%.

3.2. Study Sample

The sample was determined using the PRISMA diagram (2020), which helped in obtaining the sample systematically as it is a guide that clearly organizes the steps of selecting research and documents its sources (Page et al., 2021). The following diagram in Figure (1) clearly represents the sample:

Figure (1): PRISMA (2020) diagram showing how the sample was obtained



3.3. Research tools

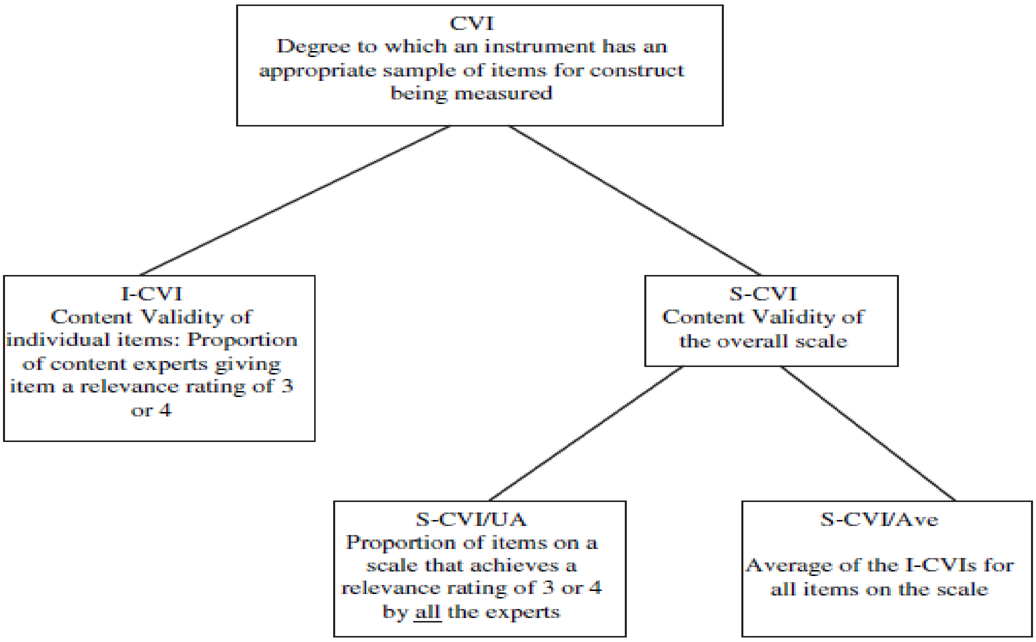
3.3.1. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria for the sample were determined after a broad review of published research relevant to the study topic. A set of criteria was extracted and formulated in accordance with the study objectives, and the literature from which they were extracted is shown in the appendix.

3.3.2. Analysis Card

The analysis card was developed after reviewing previous literature, including studies by Li and Zhang (2021); Pereira et al. (2022); Yan et al. (2023); Youde (2019), and other relevant studies. The card was designed to test the suitability of the current study's questions and objectives for analysis, thereby systematically and uniformly extracting and coding data from the research included in the study sample. Following the card's development, its validity was confirmed by presenting the initial version to 16 experts. To ensure the validity of the analysis card, we relied on the indicators shown in the following figure:

Figure 2: Content Validity Index (Polit & Beck, 2006)



After making adjustments based on the reviewers' feedback, they were asked to evaluate the items according to Lawshe (1975) scale, which consists of three levels: appropriate and necessary, appropriate but not necessary, and inappropriate and not necessary. This was done to calculate the Coefficient of Variation (CVR). According to Ayre and Scally (2014), the CVR indicates that the content validity of all items on the analysis card (the relative variance between reviewers) was within the acceptable level. The number of reviewers agreeing on the criterion (N critical) must be at least 12 for it to be accepted when there are a total of 16 reviewers. In this case, the CVR values must be at least the critical value (CVR critical = 0.05). Thus, all items were within the acceptable level. CVI (Certainty of Validity Index) indices were calculated to verify the validity of the items by asking reviewers to evaluate them using a four-point linear scale according to Ansari and Khan (2023): Needs Adjustment, Partially Suitable, Suitable, Very Suitable. Initially, I-CVI (Item-Level Content Validity Index) was calculated for each item separately. All items on the analysis card were at an acceptable level, as indicated by Polit and Beck (2006), since all items had an I-CVI > 0.78.

To measure the overall validity of the items, the Scale-Level Content Validity Index (S-CVI/Ave) was calculated using the Average Method (Average Method). The result was (I-CVI/Ave = 0.98), which is considered acceptable, as S-CVI/Ave ≤ 0.90 is considered an indicator of overall item validity (Shi, Mo, & Sun, 2012). Additionally, the S-CVI/UA (Scale-Level Content Validity Index – Universal Agreement) was calculated. The S-CVI/UA value was 0.83, which is also considered acceptable. Shi et al. (2012) considers S-CVI/UA values ≥ 0.80 to be within the acceptable range. To ensure the reliability of the analysis card, 44 studies were used, obtained after applying the analysis criteria to the pilot sample consisting of 50 studies. Reliability was calculated in two ways: First, through Inter-rater reliability, relying on three experts from Appendix (8), by calculating reliability using the Holtse equation (Mao, 2017):

$$R = \frac{3\,M}{N_1 + N_2 + N_3}$$

The reliability coefficient was calculated for each item, and the results showed that the reliability level for all items ranged between (0.77 - 1) as shown in Appendix (13), which is within the acceptable level according to Holsti (1969). The reliability coefficient for the card as a whole was also calculated, and the result represented an acceptable level of reliability:

$$R = \frac{3 \times 1124}{3 \times (29 \times 44)} = 0.88$$

3.4. Cochrane Tools for Assessing the Risk of Bias

The risk of bias in the research was assessed prior to the meta-analysis using internationally recognized tools. The researcher used Cochrane tools to ensure the quality of the included studies and enhance the credibility of the meta-analysis results.

3.4.1. RoB 2 :Revised Cochrane risk-of-bias tool for randomized trials

The RoB 2 tool was specifically developed to assess the risk of bias in randomized trials. It is characterized by its outcomes-based approach, where each outcome of the study is assessed separately, providing a more accurate and detailed evaluation (Nejadghaderi, Balibegloo, & Rezaei, 2024). It focuses on the data production process and its impact on the results, rather than solely on the general methodological aspects of the study (Minozzi et al., 2022). The RoB 2 tool comprises five key areas of bias risk.

3.4.2. ROBINS-I V2: Version 2 Risk Of Bias In Non-randomized Studies of Interventions

This tool is a powerful tool for assessing the risk of bias in non-randomized studies, in which individuals are not randomly assigned to groups. This tool was developed to meet a pressing need in scientific research, as non-randomized studies constitute a large proportion of the literature (Sterne et al., 2016). This tool is distinguished by its unique approach of comparing non-randomized studies to a hypothetical ideal randomized trial as a benchmark for detecting how far non-randomized research deviates from the optimal design that ensures the least amount of bias (Igelström et al., 2021).

4. Research Results and Discussion

This chapter presents the results of the study, which the researcher obtained through sample analysis. This analysis will answer the study questions systematically, ensuring the validity of the results. Therefore, initially, Hedges'(g) effect size was calculated to estimate the effect size of alternative assessment on non-affective learning outcomes in all the studies included in the sample. All effect sizes were then standardized and converted to Hedges'(g) effect size to provide a unified measure that allows for combining the results of different studies and arriving at an effect size representative of these studies. Hedges' effect size is considered the most accurate and reliable for meta-analysis compared to other sizes, such as Cohen's(d) effect size, which appeared repeatedly in the studied studies. Cohen's (d) effect size tends to overestimate the effect size when sample sizes are small. Therefore, the researcher converted all effect sizes to Hedges'(g) effect size because it corrects for this bias using a correction factor that depends on the sample size, thus reducing bias and yielding more reliable results (Taylor & Alanazi, 2023). This was done using (Hedges, 1981) equation:

$$g = \left(\frac{\bar{X}_1 - \bar{X}_2}{S_p}\right) \times \left(1 - \frac{3}{4\,(n_1 + n_2) - 9}\right)$$

The appendices show a Forest Plot, which provides a visual summary of the effect sizes obtained from the 253 research samples. The plot displays the effect sizes and confidence intervals for each study individually, along with the overall effect size. To determine the overall effect size of the answer to the first study question (What is the overall effect size of using alternative assessment methods on non-affective learning outcomes?) and to reveal the degree of residual variance between the studies, a classic meta-analysis was performed using a fixed-effects model, and the results are shown in the tables below.

Table 1: Results of the Classical Meta-Analysis

	Q	df	p
Omnibus test of Model Coefficients	524.971	1	< .001
Test of Residual Heterogeneity	654.158	252	< .001

Note. *p* -values are approximate.

	Q	df	p

Note. The model was estimated using Restricted ML method.

The table of results from the classical meta-analysis shows the result of the Omnibus test of Model Coefficients, which showed that Cochran's Q (df=1) = (524.971), statistically significant at the $p < 0.001$ level. This means that the overall effect size is statistically significant, i.e., the average effect size across the studies is not zero, and there is an overall effect size that can be analyzed.

Table 2: Estimates of statistical coefficients for the overall effect size

	Estimate	Standard Error	z	p
intercept	0.442	0.019	22.912	< .001

Note. Wald test.

The results of the test of residual heterogeneity show that Q (df = 252) = (654.158), which is statistically significant at the ($p < 0.001$) level. This indicates that the research in the sample was not entirely homogeneous, meaning that the effect sizes differed more than could be explained by random error alone. This justifies the use of a random effects model. Furthermore, the table of estimated statistical coefficients indicates that the overall effect size was (0.44) at ($p < 0.001$), meaning that the alternative assessment had a statistically significant positive effect.

Table 3: Estimates of residual variance and indices of variance between studies

	Estimate
τ^2	0.048
I^2 (%)	60.347
H^2	2.522

The table of estimates of residual variance and indicators of variance between studies also presents a set of indicators that estimate the residual variance between studies, where (τ^2 = (0.048), which indicates the existence of variance between effect sizes not due to random error. Also, (I^2 = (60.347%) represents the percentage of true variance between studies. The results also showed that the amount of increase of the total variance over the expected variance due to random error, H^2 = (2.522 > 1), which indicates the existence of true variance between studies, as the total variance is approximately 2.5 times greater than the variance expected by chance or random error. In general, all these indicators indicate the existence of true variance between studies in effect sizes, and this calls for an investigation into the possibility of moderators explaining at least part of this discrepancy. To answer the second question, which states: (Do the effect sizes differ depending on the type of alternative assessment used in the studies included?). a descriptive analysis of effect sizes was conducted using a box plot to illustrate their distribution according to the different types of alternative assessment. This procedure aims to provide an initial visual representation of the degree of difference or similarity between the effect sizes associated with each type of alternative assessment before conducting statistical tests.

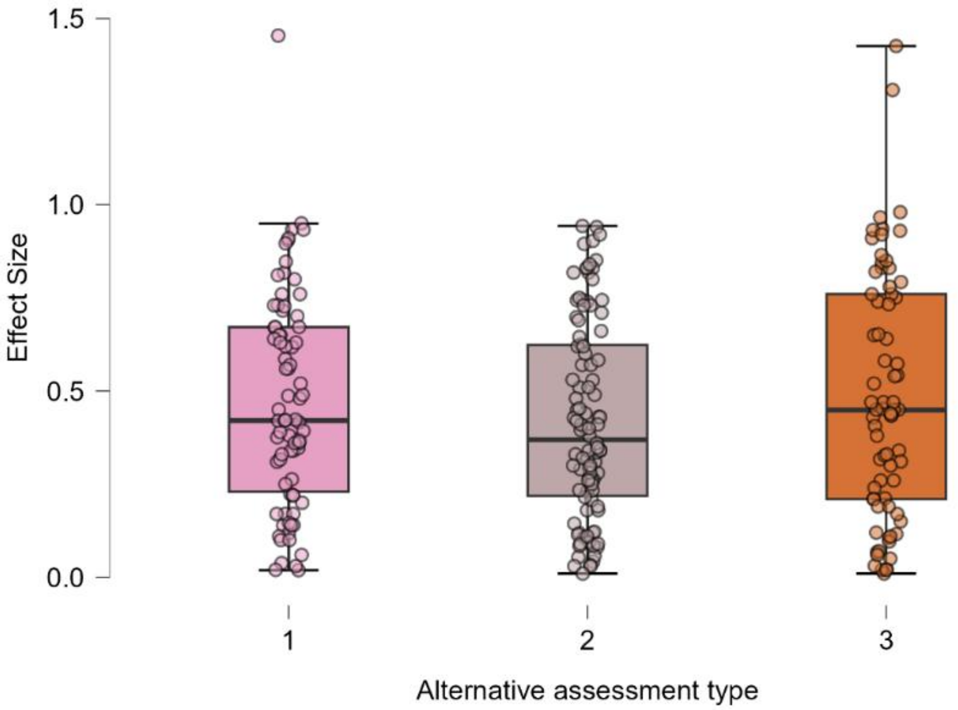
Table 4: Statistical indicators of effect sizes according to the alternative assessment type variable

	Effect Size		
	Self-Assessment (1)	**Peer Assessment (2)**	**Technology-based assessment (3)**
Valid	78	102	73
Mean	0.470	0.417	0.480
Std. Deviation	0.285	0.264	0.331

Table (4) shows the number of effect sizes available in the sample related to the alternative assessment type variable. The number of effect sizes, categorized by type, was as follows: self-assessment (78), peer assessment (102), and technology-based assessment

(73). The table also shows the mean effect sizes for each type: (0.470, 0.417, 0.480), respectively. While the mean effect sizes differed, they were relatively close.

Figure 3: Boxplot showing the distribution of effect sizes according to the alternative assessment type variable



The box plot illustrates the distribution of effect sizes according to the alternative assessment type variable. The median values for the effect sizes appear relatively similar across the three types, with the median for technology-based assessment showing a slight increase. The interquartile range (IQR), represented by the length of the boxes, exhibits some variation, indicating a diversity of effect sizes in the research. The IQR appears to be the highest for technology-based assessment. Whiskers show outliers in the self-assessment, indicating relatively high effect sizes compared to the rest of the sample. Despite the observed differences between the alternative assessment types, these observations remain preliminary and descriptive. They cannot be relied upon to determine whether there are significant differences in effect sizes between the alternative assessment types. Therefore, statistical analysis is necessary to verify the significance of these differences and to determine whether the assessment type variable acts as a moderating variable.

Table 5: Classical Meta-Analysis Results According to Alternative Assessment Type Variable

	Q	df	p
Omnibus test of Model Coefficients	2.612	2	0.271
Test of Residual Heterogeneity	646.711	250	< .001

Note. *p* -values are approximate.
Note. The model was estimated using Restricted ML method.

The results of the Omnibus test of Model Coefficients in Table (5) show that the Cochrane test value was Q(df = 2) = (2.612), which is not statistically significant at (p = 0.271). This means that there are no statistically significant differences in effect sizes between the three alternative assessment methods, and therefore this variable is not considered a moderating variable in explaining the variance between effect sizes. The results showed a statistically significant overall effect size for the research included in the meta-analysis. The overall effect size according to Hedges'(g) was (0.44), which reflects a positive effect that can be described as average according to Hedges g interpretations (Hedges, 1981). This means that the three types of alternative assessment studied together (self-assessment, peer assessment, and technology-based assessment) have a positive effect on the non-affective learning outcomes of students. The researcher believes this is because alternative assessment directly affects the student's role, preventing them from remaining passive recipients of information. Instead, it empowers them to take ownership of their learning and bear responsibility for it, thus influencing their thinking and broadening their understanding, making them more capable of identifying their weaknesses. This result is consistent with numerous meta-analyses in this field, such as Gozuyesil and Tanriseven (2017); Pellegrini (2020); Pereira et al. (2022), and Yan et al. (2022) study, which confirmed that alternative assessment, in its various forms, has a significant positive impact, ranging from moderate to high in some cases.

The results of the residual variance tests, specifically shown in Table (3), all indicate a real variance between effect sizes that exceeds what can be explained by random error. This is confirmed by several studies, such as Karaman (2021) and Youde (2019) study. The researcher believes that this variation may be due to several moderating variables that include the context and methodology of the research studied in the meta-analyses, such as the educational stage of the students, the use of criteria, the nature of the sample, student training, and other variables. Therefore what the current analysis revealed about the existence of real variation is evidence that supports the need to search for moderating factors. The results showed, when examining the type of alternative assessment used, whether it was a moderating factor, that there were slight differences between the three types (self-assessment, peer assessment, and technology-based assessment). However, these differences were not statistically significant between them, which means that the type of alternative assessment is not a moderating variable in the calculated total effect size. These three types lead to similar positive results on non-affective learning outcomes This result is consistent with several previous studies that indicated no significant differences between different alternative assessment methods in terms of their impact on achievement, such as Gozuyesil and Tanriseven (2017); Pereira et al. (2022); Yan et al. (2022), and Li et al. (2021) study. The researcher believes that the lack of differences between the three types, despite their varying methods of application, stems from the fact that these methods share similar educational characteristics and their role in influencing the student during their application in developing their ability to evaluate themselves, engage in self-reflection, and stimulate their motivation for active participation.

5. Conclusion

There is no doubt that the shift to using alternative assessment methods is an educational necessity in our time that ensures the active participation of students, according to what the effect sizes of the studied research have shown. However, the variation between effect sizes leads to a new inquiry: Are all alternative assessment methods appropriate for all situations, contexts, and student characteristics? Hence, this study recommends that researchers investigate in depth the modifying factors that actually affect the effect size, and this actually helps to determine the appropriate assessment methods for each situation and educational context. Using alternative assessment solely as a manifestation of educational development is insufficient without studying and understanding its actual impact on students. The similarity in impact sizes of the studied alternative assessment methods may indicate that their impact size is not due to the application mechanisms themselves, as they are similar in their theoretical basis and philosophical essence, which focuses on developing the student's thinking and guiding him to take responsibility for his learning and increase his engagement in learning, as much as it is related to the context and circumstances in which it is applied. Therefore, relying on alternative assessment can be considered a strategic educational lever if we move from random experiments to precise, systematic application within specific contexts. Using technology for assessment, even if only to keep pace with development, is not entirely ineffective. This research indicates that technology-based assessment does not differ in its impact on affective learning outcomes from self-assessment and peer assessment. If you, as a teacher, are truly focused on results, you can replace technology-based assessment with other methods.

References

Ahmad, S., Sultana, N., & Jamil, S. (2020). Behaviorism vs constructivism: A paradigm shift from traditional to alternative assessment techniques. *Journal of Applied Linguistics and Language Research*, *7*(2), 19-33.
Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. *Frontiers in Education*, *4*, 87. https://doi.org/10.3389/feduc.2019.00087
Ansari, M. M., & Khan, S. (2023). An In-depth Examination of Validity Assessment: Exploring Diverse Methodologies and Dimensions of Validity in Social Research Studies. *Asian Journal of Agricultural Extension, Economics & Sociology*, *41*(10), 772-782. https://doi.org/10.9734/ajaees/2023/v41i102224
Ayre, C., & Scally, A. J. (2014). Critical Values for Lawshe's Content Validity Ratio: Revisiting the Original Methods of Calculation. *Measurement and Evaluation in Counseling and Development*, *47*(1), 79-86. https://doi.org/10.1177/0748175613513808
Bechu, A. M., Roy, M. A., Jacobs, M., & Tickner, J. A. (2024). Alternatives assessment: An analysis on progress and future needs for research and practice. *Integrated Environmental Assessment and Management*, *20*(5), 1337-1354. https://doi.org/10.1002/ieam.4882
Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551-575. https://doi.org/10.1080/0969594X.2018.1441807
Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to Meta‐Analysis* (1 ed.). Wiley.
Bulut, O., & Beiting-Parrish, M. (2024). The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges. *Chinese/English Journal of Educational Measurement and Evaluation*, *5*(3). https://doi.org/10.59863/MIQL7785
Çoğaltay, N., & Karadağ, E. (2015). Introduction to Meta-Analysis. In E. Karadağ (Ed.), *Leadership and Organizational Outcomes* (pp. 19-28). Springer International Publishing.
Corbin, T., Bearman, M., Boud, D., & Dawson, P. (2025). The wicked problem of AI and assessment. *Assessment & Evaluation in Higher Education*, 1-17. https://doi.org/10.1080/02602938.2025.2553340
Daneshfar, S., & Moharami, M. (2018). Dynamic Assessment in Vygotsky's Sociocultural Theory: Origins and Main Concepts. *Journal of Language Teaching and Research*, *9*(3), 600. https://doi.org/10.17507/jltr.0903.20
Demir, M. (2021). Alternative assessment methods in primary education: Review and future directions. *Current studies in educational disciplines*, *227*.
Fang, Y., Roscoe, R. D., & McNamara, D. S. (2023). Artificial intelligence-based assessment in education. In B. Du Boulay, A. Mitrovic, & K. Yacef (Eds.), *Handbook of Artificial Intelligence in Education* (pp. 485-504). Edward Elgar Publishing.
Flavell, J. H. (2024). Metacognitive Aspects of Problem Solving. In *The Nature of Intelligence* (1 ed., pp. 231-236). Routledge.
Fleckney, P., Thompson, J., & Vaz-Serra, P. (2025). Designing effective peer assessment processes in higher education: a systematic review. *Higher Education Research & Development*, *44*(2), 386-401. https://doi.org/10.1080/07294360.2024.2407083
González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial Intelligence for Student Assessment: A Systematic Review. *Applied Sciences*, *11*(12), 5467. https://doi.org/10.3390/app11125467
Gozuyesil, E., & Tanriseven, I. (2017). A Meta-analysis of the effectiveness of alternative assessment techniques. Eurasian Journal of Educational Research. *Eurasian Journal of Educational Research*, *17*(70), 1-20. https://doi.org/10.14689/ejer.2017.70.3
Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *Journal of Educational Statistics*, *6*(2), 107-128. https://doi.org/10.3102/10769986006002107
Herman, J. L. (1992). *A practical guide to alternative assessment*. ERIC.
Holsti, O. R. (1969). Content analysis for the social sciences and humanities. *Reading. MA: Addison-Wesley (content analysis)*.
Igelström, E., Campbell, M., Craig, P., & Katikireddi, S. V. (2021). Cochrane's risk of bias tool for non-randomized studies (ROBINS-I) is frequently misapplied: A methodological systematic review. *Journal of Clinical Epidemiology*, *140*, 22-32. https://doi.org/10.1016/j.jclinepi.2021.08.022
Jacobs, G. M., & Renandya, W. A. (2019). Alternative Assessment. In *Student Centered Cooperative Learning* (pp. 61-74). Springer Singapore.

Karaman, P. (2021). The Impact of Self-assessment on Academic Performance: A Meta-analysis Study. *International Journal of Research in Education and Science*, *7*(4), 1151-1166. https://doi.org/10.46328/ijres.2344

Lawshe, C. H. (1975). A QUANTITATIVE APPROACH TO CONTENT VALIDITY[1]. *Personnel Psychology*, *28*(4), 563-575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Leppink, J. (2019). *Statistical Methods for Experimental Research in Education and Psychology*. Springer International Publishing.

Levitt, H. M. (2024). How to conduct an integrative mixed methods meta-analysis: A tutorial for the systematic review of quantitative and qualitative evidence. *Psychological Methods*. https://doi.org/10.1037/met0000675

Li, H., Bialo, J. A., Xiong, Y., Hunter, C. V., & Guo, X. (2021). The Effect of Peer Assessment on Non-Cognitive Outcomes: A Meta-Analysis. *Applied Measurement in Education*, *34*(3), 179-203. https://doi.org/10.1080/08957347.2021.1933980

Li, M., & Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing*, *38*(2), 189-218. https://doi.org/10.1177/0265532220932481

Luan, R. J., & Narayanan, R. (2024). ENHANCING STUDENT ENGAGEMENT WITH AUTHENTIC ASSESSMENT IN TECHNOLOGY-BASED PROGRAM: A MULTI-FACETED APPROACH. *International Journal of Modern Education*, *6*(21), 655-664. https://doi.org/10.35631/IJMOE.621048

Maki, P. L. (2023). *Assessing for Learning: Building a Sustainable Commitment Across the Institution* (2 ed.). Routledge.

Mao, Y. (2017). Intercoder reliability techniques: Holsti method. *The sage encyclopedia of communication research methods*, *2*, 741-743.

Maynard, R. (2024). Improving the Usefulness and Use of Meta-Analysis to Inform Policy and Practice. *Evaluation Review*, *48*(3), 515-543. https://doi.org/10.1177/0193841X241229885

Meylani, R. (2024). A Comparative Analysis of Traditional and Modern Approaches to Assessment and Evaluation in Education. *Batı Anadolu Eğitim Bilimleri Dergisi*, *15*(1), 520-555. https://doi.org/10.51460/baebd.1386737

Minozzi, S., Dwan, K., Borrelli, F., & Filippini, G. (2022). Reliability of the revised Cochrane risk-of-bias tool for randomised trials (RoB2) improved with the use of implementation instruction. *Journal of Clinical Epidemiology*, *141*, 99-105. https://doi.org/10.1016/j.jclinepi.2021.09.021

Monib, W. K., Karimi, A. Q., & Nijat, N. (2020). Effects of Alternative Assessment in EFL Classroom: A Systematic Review. *American International Journal of Education and Linguistics Research*, *3*(2), 7-18. https://doi.org/10.46545/aijelr.v3i2.152

Naka, L. (2025). Portfolio as an assessment tool: Impact on student participation and improvement in EFL learning. *International Journal of Language Testing*, *15*(2), 134-152.

Nejadghaderi, S. A., Balibegloo, M., & Rezaei, N. (2024). The Cochrane risk of bias assessment tool 2 (RoB 2) versus the original RoB: A perspective on the pros and cons. *Health Science Reports*, *7*(6), e2165. https://doi.org/10.1002/hsr2.2165

Ng, J., Szydlowski, N., Gill, M., Fusco, N., & Ruiz, K. (2024). MSR72 An Evaluation of the Rayyan Artificial Intelligence Tool for Systematic Literature Review Screening. *Value in Health*, *27*(6), S273. https://doi.org/10.1016/j.jval.2024.03.1505

Nieminen, J. H., & Boud, D. (2025). Placing authenticity at the heart of student self-assessment: an integrative review. *Teaching in Higher Education*, *30*(3), 640-662. https://doi.org/10.1080/13562517.2024.2367669

Nieminen, J. H., Haataja, E., & Cobb, P. J. (2025). From active learners to knowledge contributors: authentic assessment as a catalyst for students' epistemic agency. *Teaching in Higher Education*, *30*(4), 970-990. https://doi.org/10.1080/13562517.2024.2332252

Nisbet, I., & Shaw, S. (2024). *Educational Assessment in a Changing World: Lessons Learned and the Path Ahead* (1 ed.). Routledge.

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74-98. https://doi.org/10.1016/j.edurev.2017.08.004

Pellegrini, M. (2020). The effectiveness of peer assessment on students' performance in higher education. Evidence of an overview of meta-analyses. *Form@re - Open Journal per la formazione in rete*, *20*(1), 128-142. https://doi.org/10.13128/form-8339

Pereira, P., Heitink, M., Schildkamp, K., Veldkamp, B., & Feskens, R. (2022, 09/2022). A meta-analysis of the effect of peer feedback on academic achievement in the STEM fields of higher education. SEFI 50th Annual conference of The European Society for Engineering Education,

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health*, *29*(5), 489-497. https://doi.org/10.1002/nur.20147

Quinlan, K. M., Sellei, G., & Fiorucci, W. (2025). Educationally authentic assessment: reframing authentic assessment in relation to students' meaningful engagement. *Teaching in Higher Education*, *30*(3), 717-734. https://doi.org/10.1080/13562517.2024.2394042

Raslan, G. (2024). The Impact of the Zone of Proximal Development Concept (Scaffolding) on the Students Problem Solving Skills and Learning Outcomes. In K. Al Marri, F. A. Mir, S. A. David, & M. Al-Emran (Eds.), *BUiD Doctoral Research Conference 2023* (Vol. 473, pp. 59-66). Springer Nature Switzerland.

Seth, P., & Sankarapu, V. K. (2025). Bridging the gap in XAI-why reliable metrics matter for explainability and compliance. *arXiv preprint arXiv:2502.04695*. https://doi.org/https://doi.org/10.48550/arXiv.2502.04695

Shapiro, J. P., Gross, S. J., Shapiro, S. H., Aronow, S. L., O'Connor, J. M., Murphy, R. J., Caldwell, C. A., Brasof, M., & Myers, E. L. (2025). Power Versus Accommodation in Curriculum, Instruction, and Assessment. In *Ethical Educational Leadership in Turbulent Times* (3 ed., pp. 87-113). Routledge.

Shavelson, R. J., Zlatkin-Troitschanskaia, O., Beck, K., Schmidt, S., & Marino, J. P. (2019). Assessment of University Students' Critical Thinking: Next Generation Performance Assessment. *International Journal of Testing*, *19*(4), 337-362. https://doi.org/10.1080/15305058.2018.1543309

Shi, J., Mo, X., & Sun, Z. (2012). Content validity index in scale development. *Zhong nan da xue xue bao. Yi xue ban= Journal of Central South University. Medical sciences*, *37*(2), 152-155. https://doi.org/https://doi.org/10.3969/j.issn.1672-7347.2012.02.007

Sinaga, Y. D. K., Arliani, E., Ngala, J. C., & Agustina, N. L. I. T. (2024). Accuracy of Self-Assessment and Peer Assessment in Learning : A Systematic Literature Review. *Jurnal Paedagogy*, *11*(2), 312. https://doi.org/10.33394/jp.v11i2.9417

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., & Boutron, I. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *bmj*, *355*.

Sukumal, P. (2024). Assessing learning outcomes in Thai secondary schools: A comparative study of traditional and alternative evaluation methods. *Insights into Modern Education (i-ME)*, *1*(6), 53-66.

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, *3*, 100075. https://doi.org/10.1016/j.caeai.2022.100075

Taylor, J. M., & Alanazi, S. (2023). Cohen's and Hedges' g. *Journal of Nursing Education*, *62*(5), 316-317. https://doi.org/10.3928/01484834-20230415-02

Vallis, C. (2025). Authentic assessment in higher education: the spectre of lost futures. *Teaching in Higher Education*, *30*(3), 744-751. https://doi.org/10.1080/13562517.2024.2362217

Wakeman, S. Y., Karvonen, M., Flowers, C., & Ruhter, L. (2021). Alternate Assessments and Monitoring Student Progress in Inclusive Classrooms. In *Handbook of Effective Inclusive Elementary Schools* (2 ed., pp. 302-321). Routledge.

Wu, C.-T., & Chang, C.-W. (2025). Applications of AI in Educational Research Series (XIII): AI-Assisted Systematic Literature Review and Rayyan Application. *Journal of Education Research*(379), 119-141. https://doi.org/10.53106/168063602025110379008

Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education*, *45*(2), 224-238. https://doi.org/10.1080/02602938.2019.1629390

Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review*, *37*, 100484. https://doi.org/10.1016/j.edurev.2022.100484

Yan, Z., Wang, X., Boud, D., & Lao, H. (2023). The effect of self-assessment on academic performance and the role of explicitness: a meta-analysis. *Assessment & Evaluation in Higher Education*, *48*(1), 1-15. https://doi.org/10.1080/02602938.2021.2012644

Youde, J. J. (2019). A meta-analysis of the effects of reflective self-assessment on academic achievement in primary and secondary populations.