Vol. 20(6), pp. 101-113, June 2025
DOI: 10.5897/ERR2025.4469
Article Number: 926A63573352
ISSN: 1990-3839

http://www.academicjournals.org/ERR

ACADEMIC JOURNALS
expand your knowledge

Educational Research and Reviews

Full Length Research Paper

Explanatory variables influencing outcomes in government standardized tests at the upper secondary level: A scoping review

Eliacid Marcelo Escalante[1], Sandra Patricia Barragán Moreno[2]* and Alfredo Guzmán Rincón[3]

[1]Secretary of Education of Bogotá, University of Bogotá Jorge Tadeo Lozano, Bogotá, Colombia.
[2]Doctoral Program in Modeling in Public Policy and Management, University of Bogotá Jorge Tadeo Lozano, Bogotá, Colombia.
[3]Hub top managers, Corporación Universitaria de Asturias, Colombia.

Received 11 April, 2025; Accepted 6 June, 2025

This study aims to synthesize findings on the variables that explain the results of government standardized tests at the secondary education level in educational institutions. A scoping review methodology was applied, following the PRISMA protocol, to identify and analyze 33 relevant academic articles retrieved from major scientific databases. A bibliometric analysis showed a steady increase in publications on the topic since 2019. A keyword co-occurrence analysis grouped the literature into four thematic clusters: educational policies, assessment strategies, institutional factors, and social equity. The findings reveal that standardized test outcomes are shaped by factors endogenous to students (e.g., family and social background), to educational institutions (e.g., pedagogical models, resources, organizational strategies), and to national and international policy environments. The discussion highlights the limitations of traditional approaches that emphasize test scores in isolation and advocates for formative, inclusive evaluation practices that are sensitive to local contexts. A systemic and integrated perspective is necessary to use standardized test results as meaningful tools for improving educational quality and guiding equitable public policy decisions in education.

Key words: Pedagogical evaluation, educational quality, standardized tests, public policy, educational indicators.

INTRODUCTION

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2020) defines education as the process of facilitating learning and the acquisition of knowledge, skills, values, beliefs, and habits. Education, in the context of this study, is understood as a transformative and socially embedded process that fosters the development of cognitive, emotional, and ethical capacities in individuals, enabling them to participate meaningfully in society. It is essential to refer to key performance indicators to monitor and enhance the quality of education across all levels of educational institutions (Vinajera-Zamora et al., 2023). Consequently, a significant body of research has concentrated on higher education quality, including programme accreditation

*Corresponding author. E-mail: Sandra.barragan@utadeo.edu.co. Tel: +57 3204929701.

(Bougherira and Mayada, 2023), learning outcomes (Araujo et al., 2021; Alyasin et al., 2023), financial sustainability (*Asociación Colombiana de Universidades* [Colombian Association of Universities], 2022; Cosenz, 2022), student dropout (National Ministry of Education, 2009), academic productivity and results in standardized tests (Bogoya, 2022; Bogoya et al., 2014). In contrast, quality indicators for primary and secondary education(including upper secondary education) tend to focus on school performance (Henning et al., 2021; Díaz, 2019), academic progress (ECLAC, 2022; Cabrera et al., 2014), efficiency (Lozano, 2022), inclusion (Correa, 2016), student dropout (Herrera, 2021; Ramírez and Carcausto, 2024), institutional leadership (Goedert and Forcellini, 2023), teaching quality (Brutti and Sánchez, 2022; Onosko, 2011), and outcomes in standardized assessments (OECD, 2019; Sharif et al., 2022; Fernández et al., 2019).

Government-standardized tests results have emerged as key indicators for assessing education quality, particularly at the upper secondary level. These results not only contribute to the analysis of teaching but also inform the evaluation of public education policies. According to the World Bank (2010), a strong correlation exists between test outcomes and the factors influencing educational quality, which are crucial to bridging social inequalities. Standardized testing has been implemented globally, with numerous reports documenting their characteristics and assessment methodologies. A notable example is the Technical Report of the PISA Programme (PISA, 2006), which outlines the conceptual foundations and measurement procedures used with nearly 400,000 students across 57 countries. Similarly, the TIMSS & PIRLS (2009) technical report examined the evaluation of mathematical and scientific competencies in students in grades 4 and 8, involving 67 participants (59 countries and 8 benchmarking entities) (Mullis et al., 2015). In the Latin American context, Ramos et al. (2023) analyzed how learning evaluations have been investigated through online learning platforms as tools for standardized tests.

Given the triad of educational purposes, goals, and governmental priorities (as well as the outcomes of national and international standardized assessment) a gap in literature becomes apparent. This underscores the need to gain a deeper understanding of the various variables that influence student performance in such tests. It is widely acknowledged that individual characteristics, socioeconomic background, institutional factors, and student-specific dynamics play crucial roles. Although prior studies have explored the variables that impact standardized tests results (Logan and Burdick, 2016; Barragán and Marcelo, 2022; Abadía et al., 2023; Smith et al., 2023), there remains a need for a comprehensive synthesis that captures and analyzes these dimensions holistically. This would allow for a more complete portrayal of the current educational landscape. Equally important is the examination of strategies that have been proposed or implemented to enhance performance in standardized tests. Such analysis must be firmly grounded in empirical evidence related to the explanatory variables under investigation. Therefore, this article aims to synthesize findings on the variables that explain outcomes in government-standardized tests at the upper secondary education level in educational institutions. To guide this synthesis, the study was structured around the following research questions:

RQ1: What is the bibliometric evolution of publications addressing the outcomes of government-standardized tests applied at the end of upper secondary education as indicators of educational quality?
RQ2: Which individual, academic, socioeconomic, and institutional variables influence the results of government-standardized tests applied at the end of upper secondary education in educational institutions?
RQ3: What alternatives have been considered to improve the level of academic performance in government standardized tests applied at the end of high school education?

To address these questions, the article was organized into five parts. The first discusses the foundational components of educational quality, including its multidimensional indicators, the role of assessment practices in learning practices in monitoring learning, and the theoretical underpinnings of standardized testing as a policy and pedagogical tool. The second details the research methodology, emphasizing the use of a scoping review framework aligned with PRISMA-SCR guidelines. The third section presents the main findings, the explanatory variables influencing test outcomes, and the strategies identified for performance improvement. The fourth section offers a critical discussion of these findings based on current debates with a synthesis of the key contributions, practical implications, and identified gaps for future research.

Despite numerous studies addressing educational quality and assessment reforms independently, a significant gap remains in understanding how various educational factors—from students' socioeconomic backgrounds to institutional practices and national policies—interact with shape standardized assessment outcomes. This study seeks to bridge that gap by conducting a systematic scoping review and bibliometric analysis to identify key explanatory variables, as well as the reforms, transformations, and changes proposed in the literature. Anchoring this research into the persistent challenges of equity, curricular coherence, and effective assessment systems provides a strong rationale for synthesizing and visualizing the current state of knowledge. The findings aim to inform the development of more integrated and evidence-based education policies.

Table 1. Factors influencing the quality of education.

Factor		Characterization
Endogenous to students	Social	Demographic context, housing conditions, social stratification, environment, commuting times, transport methods.
	Family	Household composition, parental educational levels, parental occupations.
	Partner level economic	Household income, availability of resources.
	Educational models	Pedagogical approaches, teaching strategies, academic planning, assessment frameworks.
	Resources	Human: Teachers, school leaders, administrative staff, institutional partnerships.
		Technical: Technological infrastructure, teaching materials.
		Physical: Classrooms, learning environments.
Endogenous to EI	Organizational	Strategic planning, school improvement plans, community engagement, school schedules.
	Assessment outcomes	Results from internal and external standardized tests, including international assessment.
	Equity and support	Budget allocation, contract schemes, social support programs.
	Supplementary plans	School meals, transport, extracurricular activities.
	Service characteristics	Coverage, inclusiveness, infrastructure.
	School facilities	Provision of technological, physical, structural and pedagogical resources.
Endogenous to the nation	State policies	Public policies, national development plans, Ten-Year Education Plans, regional education plans.
Exogenous to the nation	International participation	Multinational agreements, alignment with global educational strategies.

Source: Elaborated by the authors based on UNESCO (2005), UNICEF (2020), Barba (2018), OCDE (2019) and Castañeda (2021).

LITERATURE REVIEW

To synthesize the relevant aspects of standardized testing, it is essential to examine three core dimensions in which such assessments are embedded: the quality of education, the quality of assessment, and the conceptualization of evaluation in the context of standardized tests.

Quality of education

Educational quality is framed within a broad spectrum of approaches, theories, and perspectives, underpinned by diverse ideologies that shape contemporary educational development (UNESCO, 2005). UNICEF (2020) links educational quality to both internal (endogenous) and external (exogenous) factors associated with the student, educational institutions, and the broader national context. Yao and Lin (2023) suggest that educational quality hinges on the alignment between state policies, the characteristics of the educational provision, and user needs. The Organization for Economic Cooperation and Development (OECD) (2019), meanwhile, associates educational quality with the population´s academic attainment, economic development, and national income level. Table 1 presents the principal factors influencing educational quality, based on the literature reviewed. Table 1 highlights the multidimensional nature of educational quality, showing how student, institutional, national, and international factors interact. Social and economic conditions, educational models, resource availability, and policy frameworks all contribute to learning outcomes. Notably, institutional variables (such as teaching practices, infrastructure, and assessment systems) play a key role in mediating the impact of external and personal factors. This integrated perspective reinforces the importance of coordinated strategies across all levels to improve educational performance and equity. Castro and Ruiz (2019) affirm that standardized tests are valuable tools for characterizing student populations. Furthermore, the OECD (2019) and the World Bank (2008) have highlighted how the dissemination of test results, disaggregated by demographic and contextual variables, supports country rankings and provides a strategic basis for designing national and international education policies.

Quality of assessment

Assessment in education is understood as a continuous

and personalized process within the teaching-learning dynamic (Merke et al., 2021). Its purpose is to monitor student progress and implement remedial or compensation measures, when necessary (Castañeda, 2021), thereby ensuring the achievement of learning objectives (Gardner, 2019; Meeter, 2022).

Effective assessment requires well-defined criteria, components, and standards for design and implementation. Logan and Burdick (2016) argue that assessment frameworks incorporating diverse strategies and methodologies foster greater educational equity. Thus, assessment practices should aim to transform teaching processes, including complementary activities that facilitate timely student progress (Jurist et al., 2016). Araujo et al. (2021) suggest that a quality assessment system should be qualitative in nature, incorporating clear goals, objectives, methodologies, and pedagogical environments to support effective teaching and learning (Barragán and Marcelo, 2023).

Barragán and Marcelo (2022) highlight the importance of collaborative and interdisciplinary approaches in the evolution assessment. This includes curricular alignment across school levels, early and ongoing exposure to assessment practices and the integration of technology to support students' cognitive development. Likewise, Araujo et al. (2021) point out that improvements in assessment quality require greater inclusion and progressive implementation.

Yao and Lin (2023) further contend that quality must be built on coherent concepts, principles, and norms aligned with the educational system's overarching goals.

Ultimately, assessment outcomes are reflected in students' competencies in interpretation, argumentation, and problem-solving (McCulloch et al., 2013). The evaluation process also entails formative and contextual dimensions fostering meaningful teacher-student interactions and supporting the construction of knowledge (Beartman et al., 2007; Jiménez, 2016). In this vein, Bogoya (2006) conceptualizes assessment as the recognition and development of intellectual abilities, while Castañeda (2021) asserts that evaluation extends beyond the acquisition of knowledge to include cognitive and metacognitive growth.

Conceptualization of evaluation through standardized tests

Standardized tests are widely employed as evaluative tools by various countries to gauge educational outcomes (OECD, 2019). Each country defines the areas to be assessed, test duration, item types, and scoring methods based on its own curricular framework (Barragán and Marcelo, 2022). According to Baker (2001), this evaluative model enables the monitoring of student performance and core competencies through observable indicators, often guided by the Item Response Theory (IRT).

The results of standardized assessments serve as reference points for benchmarking educational institutions on national scales (Meeter, 2022). These outcomes also provide a more objective indicator of educational quality (Bogoya et al., 2014). In the international arena, test results from PISA, TIMSS, SERCE, and LLECE enable cross-country comparisons of student competencies, both general and subject-specific (IEA & PIRLS, 2021; ICFES, 2020).

Fernández et al. (2019) note that such assessments shape global education policy discourses and influence national systems. However, the World Bank (2008) cautions that standardized tests should not be regarded as the sole measure of educational quality. Internal school processes and pedagogical practices must also be considered.

METHODOLOGY

In alignment with the objective of this article, a scoping review was conducted as the principal methodological approach, following the typology proposed by Grant and Booth (2009). Given the complexity of the educational phenomena under study and the diversity of variables involved, a scoping review was deemed appropriate to map the breadth of existing research. This review focused on research related to outcomes from government-standardized tests at the level of upper secondary education within educational institutions. The scoping review adhered to the Preferred Reporting Items for Scoping Reviews (PRISMA-ScR) framework, ensuring transparent documentation from conceptualization, methodology, process, and results (PRISMA, 2020).

Scoping reviews are particularly suited to mapping key concepts across a broad topic by offering comprehensive coverage of the literature (Chambergo et al., 2021). This methodology supports a rigorous and inclusive synthesis process and is applicable to both published and emerging studies. It permits qualitative synthesis, which may include narrative summaries and tabular representations. Beyond documenting the volume of available literature, scoping reviews also assess the quality and the key characteristics of the included studies (Grant and Booth, 2009). As noted by Campell et al. (2023), this type of review compiles findings from a wide range of methodologies and variables, offering a holistic perspective on the research landscape.

Protocol and registration

The review protocol was developed and registered in accordance with the parameters and specifications for scoping reviews outlined by the International Platform of Registered Systematic Review and Meta-analysis Protocols (Inplasy, 2023). The registration was assigned to the code INPLASY2023120017 and the DOI 10.37766/inplasy2023.12.0017.

Eligibility and exclusion criteria

The eligibility criteria for the review were as follows: Peer-reviewed academic journal articles; Written in either English or Spanish; Published between 2013 and 2023; Focused on the outcomes of government standardized-tests administered at the end of upper secondary education; and national or state-level studies. Studies were excluded if they: Did not explicitly link standardized tests

Table 2. Documents excluded at the screened stage.

Exclusion criterion	Document	Percentage of total
Not academic journal articles	11	2
Published before 2013	41	9
Not related to student outcomes in upper secondary education.	175	39
Not of national or state-wide scope	89	20
Did not address standardized test results as indicators of educational quality.	53	11
Did not analyze influencing variables	35	8

Source: Elaborated by the authors.

outcomes with indicators of educational quality; Failed to analyze how individual, socioeconomic, academic, or institutional variables influenced; Were not peer-reviewed journal articles; Were published before 2023: Focused on levels of education other than upper secondary; Had a local or institutional scope.

Information sources

Two electronic databases were selected to ensure both breadth and relevance: SCOPUS, due to its interdisciplinary reach and extensive coverage of scholarly journals across disciplines (Baquee et al., 2023), and Education Resources Information Center (ERIC), as a discipline-specific database focused on educational research (Ahmadi et al., 2023).

Search strategy

The search was conducted on October 2, 2023, using a Boolean logic-based search equation incorporating truncation symbols to capture terms variations (Southern Adventist University McKee Library, 2023):

("Standardized test$" OR "national exam$" OR "public exam$" OR "standard assessment test$") AND ("middle school$" OR "Upper secondary education" OR "high school$") AND NOT ("higher education" OR "university$" OR "tertiary education" OR "college$" OR "Parent education" OR "Parenting education" OR "Family education" OR "Special school$" OR "home school$").

This equation was tailored to the search functionalities of each database. The terms were derived from previously published literature (Sánchez et al., 2023), ensuring theoretical relevance and terminological accuracy The initial search yielded 205 records in ERIC and 240 in SCOPUS, totaling 445 documents ($n_1 = 445$).

Selection of evidence sources

Duplicate entries ($n = 8$) were removed using Zotero reference management software. The remaining 437 documents ($n_2 = 437$) were screened by title, abstract, and keywords to determine alignment with the study's research questions. Table 2 summarizes the exclusion criteria applied at the screening stage.

Table 2 summarizes the reasons for excluding documents during the screening phase of the scoping review. A total of 402 records (90%) were discarded based on predefined criteria, including publication type, date, scope, and relevance to the research questions. The most common exclusion reason was the lack of focus on students' outcomes at the end of upper secondary education, followed by studies without a national state-level scope, or those not addressing standardized test as indicators of educational quality. After applying these criteria, 33 documents ($n_3 = 33$) were retained for full-text review and included in the bibliometric and thematic analyses.

Selection of evidence and data extraction

The selected articles were reviewed in full by the research team, who
independently extracted data into a structured spreadsheet. The extracted elements included: Reference code, title, authors, year of publication, name of the journal, methodology, individual, socioeconomic, academic, and institutional variables (drawn from findings, not theoretical frameworks or introductions), and proposed alternatives for improving student performance in standardized tests. A PRISMA flow diagram was developed to provide a visual summary of the selection process (Figure 1).

Figure 1 presents the PRISMA flow diagram, outlining the selection process of the documents included in the scoping review. It details each phase (from identification and screening to eligibility and final inclusion) ensuring transparency in the methodological process. Out of 445 records initially retrieved, 33 studies met all inclusion criteria and were analyzed in full.

Synthesis of information

To address RQ1, a co-occurrence analysis of keywords appearing in at least five articles was performed (threshold) (Al Husaeni, 2023). This allowed for the identification of semantic clusters and thematic groupings within the literature. For RQ2, a qualitative synthesis of the explanatory variables identified in the selected articles was conducted, and the results were summarized in Figure 4. In relation to RQ3, the alternative strategies for improving outcomes in standardized test were analyzed and categorized, as presented in Figure 5.

RESULTS

RQ1: What is the bibliometric evolution of publications addressing the outcomes of government-standardized tests applied at the end of upper secondary education as indicators of educational quality?

To address this question, a bibliometric analysis was performed on the final sample of 33 academic articles. Figure 2 presents the overlay visualization that was developed, identifying a total of 52 items (keywords) grouped into four thematic clusters, comprising 15, 14,

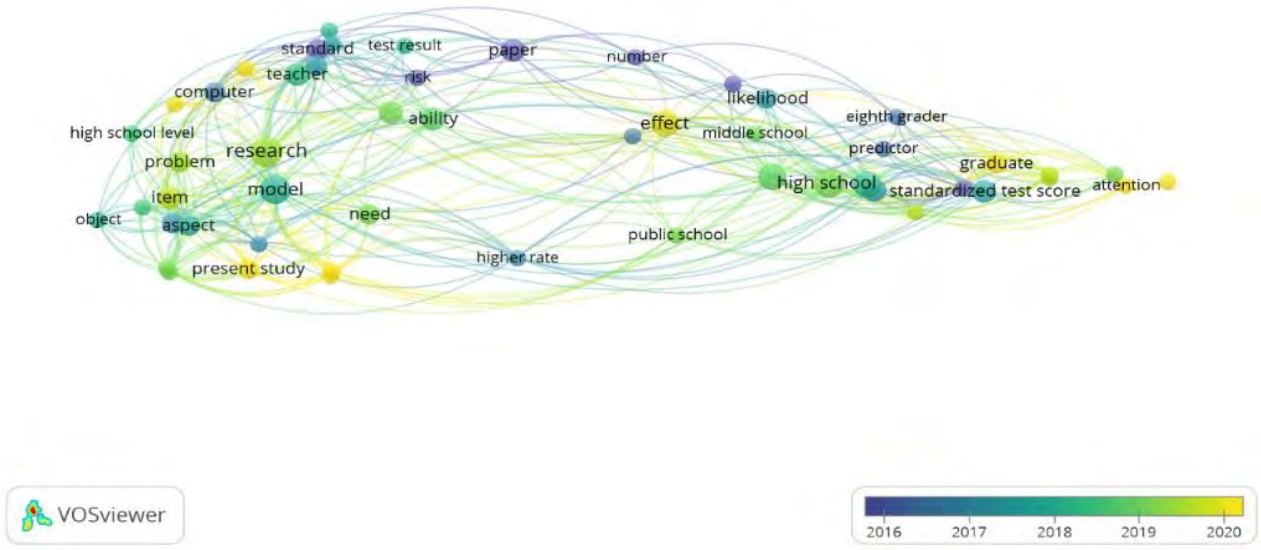

Figure 2. Co-occurrence visualization network.

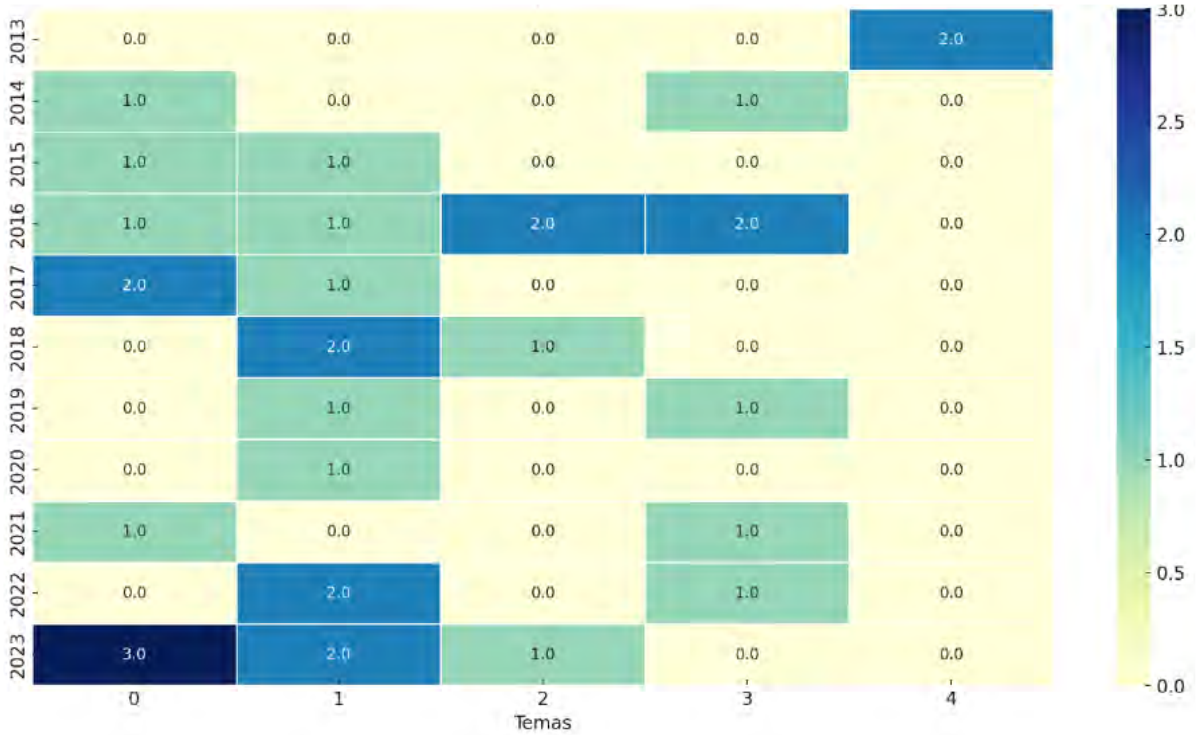

Figure 3. Heat map for years of publication and topics analyzed.

and 9 items, respectively. The ten most frequently appearing terms in article titles were: High School (100%), School (75%), Test (62.5%), followed by Math, Teachers, Student, Science, Performance, Standardized, English (each with 37.5%). This frequency underscores a pronounced scholarly focus on performance in educational settings, particularly within core subject domains among specific student groups. Publications trends by year revealed two peak periods: 2016 and 2023, each featuring six articles. These peaks suggest periodic increases in interest related to educational policy, assessment reform, or international comparison metrics.

Figure 3 presents a heat map outlining topic frequency across the publication years. This visualization enables the identification of five dominant thematic categories: 1)

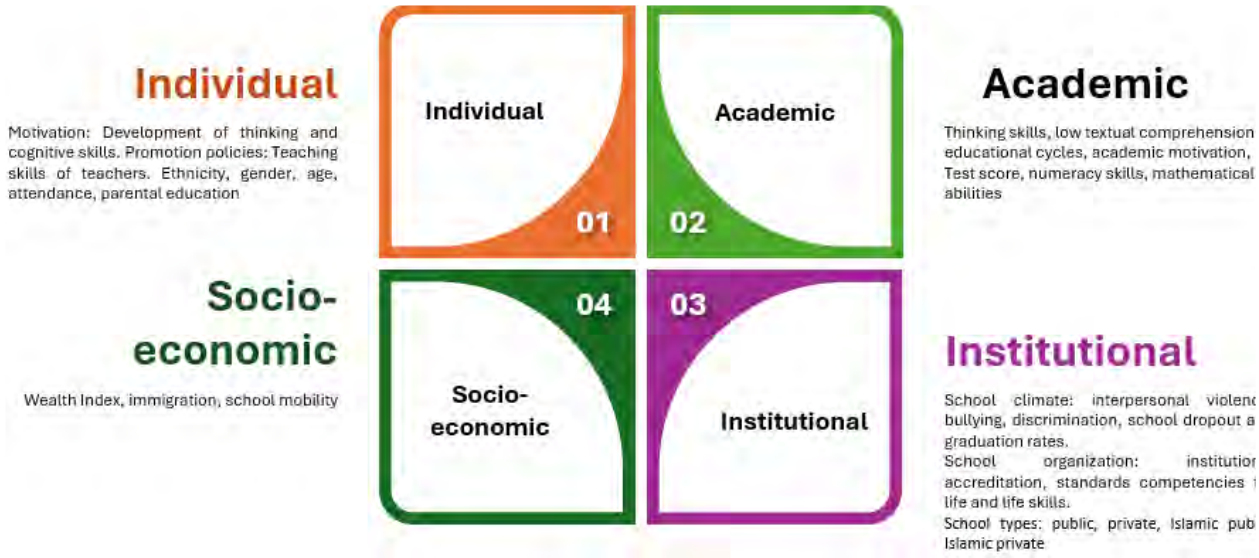

Figure 4. Explanatory variables of the results.

School education and examinations; 2) Evaluation and cognitive skills; 3) Subject-specific education and classroom dynamics; 4) National education systems and specialized topics (e.g. forestry, American education); and 5) Mathematics and cognitive educational constructs. These clusters reveal both consistency and evolution in the academic discourse on standardized testing, reflecting a dynamic and interdisciplinary research landscape.

RQ2: Which individual, academic, socioeconomic, and institutional variables influence the results of government-standardized tests applied at the end of upper secondary education in educational institutions?

The analysis of the selected articles identified a diverse range of explanatory variables, grouped into four main categories (Figure 4).

Individual variables

Several studies noted that students frequently demonstrate underdeveloped analytical and representational skills, with low levels of concentration and observation (Wiyarsi et al., 2019). Deficits in attention have also been flagged as a critical barrier to performance (Huang et al., 2022). Other works highlight the role of non-cognitive skills, such as motivation, which significantly influence student outcomes, particularly when intrinsic and extrinsic motivational factors are diminished (Smith et al., 2023; Egalite et al., 2015). Eren et al. (2017) stress that demographic attributes such as ethnicity, sex, date of birth, year they took the test, immigrant status, socioeconomic classification (e.g. Free and Reduced-Price Lunch eligibility) and prior achievement are often overlooked in policy but crucial in explaining variation in test performance.

Socioeconomic variables

Social conditions, immigration, mobility, and broader socioeconomic inequalities were consistently associated with performance outcomes. Franklin and Trouard (2016) identified age, poverty, attendance, gender, and prior test scores as key predictors. Similarly, Iglesias et al. (2023) emphasized the significance of gender, ethnicity (particularly Latino English-language learners), and school nutrition programs. Uretsky and Henneberger (2023) included drop out history, and disciplinary status (e.g. suspension, expulsion) as contributing factors. Notably, Eren et al. (2017) critiqued policies that rely solely on standardized scores for promoting decisions, as these may neglect important background information.

Academic variables

Analytical thinking, critical reading, and science-literacy integration emerged as core academic variables (Smith, 2018). Wiyarsi et al. (2019) linked higher order thinking skills to improve outcomes, while Levine et al. (2023) noted a problematic emphasis on decontextualized text interpretation. Uretsky and Henneberger (2023) found that decoding isolated or pseudowords led to lower comprehension scores. Marlatt (2018) and Deanna (2016) further indicated that student disengagement with

Figure 5. Alternatives for improving results in standardized tests.

reading tasks and classroom experiences negatively impacted outcomes.

STEM-related competencies were highlights as key performance indicators (Lee and Shibley, 2015; Chiua et al., 2013), as was the importance of timely educational progression (Uretsky and Henneberger, 2023). Teacher quality also featured prominently: Deanna (2016) underscored the need for educators to understand students' learning histories, while Egalite et al. (2015) observed a deficit in tools for assessing students' non-cognitive development. Purnomo et al. (2022) linked underperformance to inadequate teacher training in test construction and numeracy. Pollio and Hochbein (2015) suggested evolving from traditional grading practices to meeting the variety of stakeholder expectations.

Institutional variables

These variables were categorized into main areas: social climate and compliance with national standards. Bravo et al. (2020) identified interpersonal violence among adolescent's pervasive issue affecting student focus and performance.

Harris et al. (2015) and Patton and Cherng (2018) drew attention to the intersectionality of race, social class, neighborhood, and school quality, asserting that discrimination and social isolation can foster academic disengagement and dropout. From a policy perspective, Ramadhana et al. (2021) advocated for institutional alignment with accreditation standards and national competencies as a driver of improved performance in standardized testing.

RQ3: What alternatives have been considered to improve the level of academic performance in government standardized tests applied at the end of high school education?

Based on the synthesis of the reviewed literature, four broad categories of intervention emerged (Figure 5).

Evaluation reform

The analysis revealed that multiple studies emphasized the value of diversified assessment formats to better capture students' learning process. Patton and Cherng (2018) reported positive outcomes from the use of portfolios and mixed-method evaluations. Bravo et al. (2020) noted that standardized test results, when used diagnostically, supported curriculum adjustments and targeted interventions. Davis (2016) found that reducing the weight of standardized test in advancement decisions contributed to decreased stress and promote fairness.

Learning strategies

Findings from the review studies suggest that students-centered and concept-driven instructional approaches are associated with improved performance in standardized tests. Peñaranda et al. (2014) and Purnomo et al. (2022) reported that analytical reasoning and competence-based methodologies fostered deeper understanding. Deanna (2016) highlighted the benefits of contextualized reading strategies, while Martinez and Esquivel (2017)

documented improved reading comprehension through the integration of multimedia learning tools.

Motivation enhancement

Several studies underscored the impact of motivation on test performance. Smith et al. (2023) and Huang et al. (2022) identified both intrinsic and extrinsic motivators as key factors including physical activity to enhance attention and emotional regulation and eye-tracking technologies to support engagement. Alternative assessment formats, such as those reflecting students' thought processes (Smith, 2018) and competency-based grading systems (Pollio and Hochbein, 2015), were also linked to higher motivation and achievement. Calderón et al. (2020) further emphasized the importance of formative and culturally responsive evaluation practices.

Conceptual and cognitive development

The review identified that several studies reported positive outcomes from shifting away from traditional teaching models toward approaches that promote conceptual flexibility and critical thinking (Wiyarsi et al., 2019; Iglesias et al., 2023).

Additionally, Bravo et al. (2020) highlighted the effectiveness of construct-specific diagnostic assessment in guiding curriculum management and informing targeted educational interventions.

DISCUSSION

Numerous countries (including Colombia, the United States, Chile, Argentina, Mexico, Indonesia, Ecuador, Peru, Spain, and Japan) administer national standardized tests. These assessments are developed in accordance with national competency standards and evaluation criteria, as well as transnational benchmarks for countries that participate in initiatives led by organizations such as the OECD. Over time, these tests have evolved under the influence of both national and international educational public policies (OECD, 2019).

However, the outcomes of such standardized tests do not always reflect the efforts undertaken by educational institutions, educators, or governments, nor do they fully account for the impact of public funding and targeted interventions. This discrepancy has been acknowledged and analyzed by several scholar and policy observers (Abadía et al., 2023; Barragán and Marcelo, 2022).

The explanatory variables associated with student performance on standardized assessment (spanning individual, academic, socioeconomic, and institutional dimensions) highlight the diverse realities within educational institutions.

These variables reveal misalignments between educational goals, curriculum content, institutional practices, and the design of standardized assessment. Such incongruities can negatively influence outcomes and mask underlying challenges in the teaching-learning process.

The alternatives implemented across various international contexts to improve performance on standard assessment can be categorized into four key areas: learning strategies, conceptual and cognitive development, motivation, and evaluation reform. These strategies demonstrate an attempt to respond to students' academic and socio-emotional needs, although inconsistencies remain between pedagogical and institutional practices and the demands placed on students during high-stakes assessment periods.

The findings of this scoping review highlight the multifactorial nature of performance in standardized government tests at the upper secondary level, pointing to the interplay between individual, academic, socioeconomic, and institutional variables. This complexity suggests that educational outcomes cannot be solely attributed to student effort or teacher quality but are deeply embedded within broader structural and contextual factors. For instance, institutional practices such as pedagogical models, access to resources, and diagnostic use assessment data have a critical role in shaping test results. Similarly, the socioeconomic environment influences not only students' academic preparation but also their capacity to engage meaningfully with standardized evaluations. These insights underscore the need for equity focused educational strategies that move beyond one-size-fits-all approaches and instead respond to the diverse realities faced by students and schools.

The implications of these findings extend to educational policy and assessment reform. The international evidence reviewed supports the redefinition of evaluation practices shifting from high stakes testing models to more holistic, formative, and culturally responsive frameworks. Implementing diversified assessment and contextualized pedagogical strategies can enhance academic outcomes, student motivation and well-being. Furthermore, positioning standardized test results as one element within a broader quality assurance system (rather than as the sole indicator of performance) could promote more balanced accountability policies. For education systems, particularly in Latin America, these findings call for a critical reexamination of how standardized tests are integrated into policy agendas, urging greater alignment between assessment, curriculum, and support mechanism tailored to institutional and community needs.

While standardized tests results are often interpreted as proxies for educational quality, this perspective is not universally accepted. Eren et al. (2017) argue that policies which promote students solely based on their

performance in standardized tests neglect the individual variability among learners and the contextual factors that affect their outcomes. Similarly, Ross et al. (2023) contend that relying on standardized assessments as the primary metric of school quality fails to acknowledge the multifaceted efforts made by teachers (particularly in disadvantaged context) when results do not meet expectations.

Considering this, Barragán and Marcelo (2023) recommended strengthening institutional leadership in curriculum implementation and emphasizing knowledge construction across diverse educational settings, rather than relying exclusively on numerical test scores. Echoing this perspective, Ramírez and Carcausto (2024) caution against using average scores as the sole basis for positioning educational institutions in rankings, as such metrics may conceal nuanced variations in educational quality and equity.

Conclusions

The participation and results of countries in international standardized testing initiatives have become a focal point for educational policy and research, particularly for organizations such as the OECD and the World Bank. Standardized tests have evolved to reflect not only conceptual and cognitive dimensions but also policy, regulatory, and developmental aspects of educational systems. They have transformed the way learning is assessed, shifting from traditional memorization-based models to frameworks that emphasize analytical and critical thinking. This scoping review revealed that student performance in government-standardized tests at the upper secondary level is shaped by a complex interplay of individual, academic, socioeconomic, and institutional factors. Rather than viewing test outcomes in isolation, the findings support a more holistic understanding of educational quality that integrates context-sensitive variables. The review also identified key areas, such as assessment practices, motivation, and curriculum design, where improvements can significantly enhance students' outcomes. Participation of nations in international standardized tests has enabled cross-country comparisons, strengthened educational governance, and facilitated the exchange of successful methodologies. However, it has also revealed inconsistencies between the design of assessments and the contextual realities of students and institutions, particularly in settings with inequality or limited resources.

Based on the evidence analyzed, it is recommended that educational institutions prioritize diagnostic and formative assessment approaches that inform instructional planning and address students' diverse needs. Teachers should be supported in developing competence-based and student-centered methodologies that promote critical thinking. Future research should explore the impact of these strategies longitudinally and across different educational systems to strengthen evidence-based interventions. The scoping review identified that strategies for improvement, such as differentiated assessments, motivation-enhancing practices, and conceptual development tools, have proven effective in multiple contexts, but their implementation remains uneven. Therefore, it is crucial to adapt these strategies to local needs and ensure equitable access to quality teaching and learning conditions. The results of this review underscore the need for public policies that go beyond test-based accountability frameworks and instead promote equity, inclusiveness, and institutional strengthening. Standardized test results should be interpreted as part of a broader set of indicators that reflect educational quality in context. Policymakers are encouraged to align evaluation systems with curricular objectives and institutional capacities, ensuring that assessments serve both educational improvement and social justice goals.

RECOMMENDATIONS

Based on the findings of this study, several recommendations are proposed to guide future educational policies and institutional practices: First, it is essential to utilize standardized test results diagnostically to inform curricular planning and identify specific areas requiring pedagogical support. This approach enables targeted interventions aligned with students' learning needs.

Second, diversifying assessment formats through the inclusion of mixed methods, portfolios, and contextualized evaluations can provide a more comprehensive and equitable understanding of students' academic performance, especially in heterogeneous learning environments.

Third, it is advisable to reconsider the weight assigned to standardized assessments in student progression decisions. Reducing their influence may help mitigate stress and avoid reinforcing structural inequalities, particularly among disadvantaged populations.

Fourth, the transition from traditional knowledge transmission models to student-centered and competence-based instruction should be prioritized. Educational strategies that emphasize critical thinking, analytical reasoning, and conceptual flexibility can contribute more effectively to long-term learning outcomes.

Fifth, educational policies should be designed as integrated and evidence-based frameworks that reflect the interplay between institutional practices, socioeconomic contexts, and national evaluation systems. This systemic perspective is crucial to enhance the effectiveness, equity, and coherence of educational reforms.

Finally, addressing persistent challenges such as curricular fragmentation and limited equity requires strengthening institutional capacities and aligning reforms with inclusive, data-informed, and context-sensitive educational goals.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

REFERENCES

Abadía L, Gómez S, Juanita C (2023). Gone with the pandemic: How did Covid-19 affect the academic performance of Colombian students? International Journal of Educational Development 100:1-11.

Ahmadi G, Mohammadi A, Asadzandi S, Shah M, Mojtahedzadeh R (2023). What are the indicators of student engagement in learning management systems? A systematized review of the literature. International Review of Research in Open and Distributed Learning 24(1):117-136.

Al Husaeni DN (2023). Bibliometric analysis of special needs education keyword using VOSviewer indexed by Google Scholar. Indonesian Journal of Community and Special Needs Education 3(1):1-10.

Alyasin A, Nassar R, El Hajj M, Harb H (2023). Assessing learning outcomes in higher education: From practice to systematization. TEM Journal 13(3):1593-1604.

Araujo C, Martínez S, Pérez MY (2021). Do larger school grants improve educational attainment? Evidence from urban Mexico. Journal of Development Effectiveness 13(4):405-423.

Baker FB (2001). The basics of item response theory. https://www.ime.unicamp.br/~cnaber/Baker_Book.pdf

Baquee A, Rahaman MS, Sevukan D (2023). A bibliometric review of academic social networking sites (ASNSs) in scholarly communication: A scientific mapping based on Scopus database. International Journal of Information Science and Management 21(3):286-309.

Barba B (2018). La calidad de la educación: Los términos de su ecuación. Revista Mexicana de Investigación Educativa 23(76):963-979.

Barragán SP, Marcelo E (2022). Inclusive education, perspective from evaluation: Evolution in Colombian public policies. Revista Internacional de Humanidades 13(2):1-22.

Barragán S, Marcelo E (2023). Results of standardized government tests: An educational quality indicator. Frontiers in Education 8:1288640.

Beartman L, Bastiaens T, Kirschner P, Van der Vleuten C (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks. Educational Research Review 2(2):114-129.

Bogoya D (2006). Evaluación educativa en Colombia [Educational evaluation in Colombia]. Seminario Internacional sobre Evaluación Cartagena Colombia. https://drive.google.com/file/d/13shNZ7UGMl9oPWjwwYe3xUVti2V5NSz3/view

Bogoya D (2022). Elementos de calidad de la educación superior en Colombia: Caso de estudio – instituciones universitarias. Benchmarking de Universidades Colombianas 2020. https://ascofade.co/wp-content/uploads/2021/06/Benchmarking-de-universidades-colombianas-2020.pdf

Bogoya D, Barragán S, Contento M, Ocaña A (2014). Calibration of assessment instruments – classification in mathematics at the Universidad Jorge Tadeo Lozano. Revista Complutense de Educación 25(2):501-519.

Bougherira M, Mayada H (2023). Impact of academic accreditation on teaching and learning: Faculty members' perceptions. Journal of Further and Higher Education 47(2):167-181.

Bravo M, Miranda E, Miranda H (2020). Psychometric analysis of a school social climate scale: Input elements for the investigation and promotion. Frontiers in Psychology 11:605326.

Brutti Z, Sánchez F (2022). Turning around teacher quality in Latin America: Renewed. Economic Analysis and Policy 73: 62-93.

Cabrera N, Milem J, Roland JO (2014). Missing the (student achievement) forest for all the (political) trees: Empiricism and the Mexican American studies controversy in Tucson. American Educational Research Journal 51(6):1084-1118.

Calderón C, Styck K, Vega D, Kranzler J (2020). Evaluating cultural and linguistic load of IQ scores for English language learners. International Journal of School & Educational Psychology pp. 1-17.

Campell F, Tricco A, Munn Z, Pollock D, Khalil H (2023). Mapping reviews, scoping reviews, and evidence and gap maps (EGMs): The same but different—The "big picture" review family. Systematic Reviews 12:45.

Castañeda L (2021). Educational design research of a summative evaluation task in transmedia form for initial teacher's preservice training. RIED. Revista Iberoamericana de Educación a Distancia 24(2):203-224.

Castro AM, Ruíz LJ (2019). High school and higher education in Colombia seen from the Saber tests perspective. Praxis & Saber 10(24):341-346.

Chambergo D, Díaz M, Benites V (2021). Revisiones de alcance, revisiones paraguas y síntesis enfocada en revisión de mapas: Aspectos metodológicos y aplicaciones. Revista Peruana de Medicina Experimental y Salud Pública 38(1):136-142.

Chiua J, Malcolm P, Hecht D, DeJaegher C, Pan E, Bradley M (2013). WISEngineering: Supporting precollege engineering design and mathematical understanding. Computers & Education 67:142-155.

Colombian Association of Universities (2022). Analysis of the financial sustainability of higher education in Colombia. ASCUN – Asociación Colombiana de Universidades. https://ascun.org.co/documentos/analisis-de-la-sostenibilidad-financiera-de-la-educacion-superior-en-colombia/

Correa J (2016). Academic performance and gender differences in Colombia: An analysis based on the TIMMS 2007 examination results. Sociedad y Economía 30:15-42.

Cosenz F (2022). Managing sustainable performance and governance in higher education institutions: A dynamic performance management approach. Springer.

Davis E (2016). How to write science questions that are easy for people and hard for computers. Association for the Advancement of Artificial Intelligence 37(1):13-22.

Deanna M (2016). Notes and talk: An examination of a long-term English learner reading-to-learn in a high school biology classroom. Language and Education 30(3):235-251.

Díaz JE (2019). Análisis de los resultados del Índice Sintético de Calidad Educativa (ISCE) 2017 de las Secretarías de Educación certificadas de Colombia. Revista Educación y Desarrollo Social pp. 334-344.

Economic survey of Latin America and the Caribbean (ECLAC) (2022). Trends and challenges of investing for a sustainable and inclusive recovery. LC/PUB.2022/9-P/Rev.1. https://www.cepal.org/en/publications/48078-economic-survey-latin-america-and-caribbean-2022-trends-and-challenges-investing

Egalite A, Mills J, Greene J (2015). The softer side of learning: Measuring students' non-cognitive skills. SAGE Open 19(1):1-14.

Eren O, Depew B, Barnes S (2017). Test-based promotion policies, dropping out, and juvenile crime. Journal of Public Economics 9(31).

Fernández R, Woitschach P, Álvarez M, González A, Cuesta M, Muñiz J (2019). Homework and academic achievement in Latin America: A multilevel approach. Frontiers in Psychology 10:95.

Franklin B, Trouard S (2016). Comparing dropout predictors for two state-level panels using Grade 6 and Grade 8 data. The Journal of Educational Research 109(6):631-639.

Gardner H (2019). Multiple intelligences: The theory in practice. https://psicologiaymente.com/inteligencia/teoria-inteligencias-multiples-gardner

Goedert F, Forcellini F (2023). Self-assessment in higher education institutions: Literature analysis and research opportunities. Education and Research 49:e248924.

Grant M, Booth A (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. Health Information and Libraries Journal 26:91-108.
Harris A, Gundersen H, Andreassen PM, Thun E, Bjorvatn B, Ståle P (2015). Restricted use of electronic media, sleep, performance, and mood in high school. Sleep Health 1:314-321.
Henning C, Summerfeldt L, Parker J (2021). ADHD and academic success in university students: The important role of impaired attention. Journal of Attention Disorders 26(1).
Herrera D (2021). El modelo de la alternancia y la desigualdad educativa territorial en la educación en Colombia. Revista Internacional de Pedagogía e Innovación Educativa 1(2):61-86.
Huang J, Huang H, Chang B, Ho J (2022). Evaluating the effectiveness of physical exercise in improving standardized testing performances through attention indices. Brain and Behavior 12:e2800.
ICFES (2020). Informe nacional de resultados para Colombia: PISA 2018. [No disponible en línea públicamente – enlace local]
IEA & PIRLS (2021). Progress in International Reading Literacy Study 2021. Available at: https://www.iea.nl/studies/iea/pirls/2021
Iglesias V, Carriedo N, Rodríguez O, Pérez L (2023). Executive functioning skills and (low) math achievement in primary and secondary school. Journal of Experimental Child Psychology 235:105715.
Inplasy (2023). International Platform of Registered Systematic Review and Meta-analysis Protocols. Available at: https://inplasy.com/
Jiménez J (2016). El papel de la evaluación a gran escala como política de rendición de cuentas en el sistema educativo mexicano. Revista Iberoamericana de Evaluación Educativa 9(1):109-126.
Jurist A, Jia Y, Marco L, Gess J, Pasquale M (2016). Science specialists or classroom teachers: Who should teach elementary science? Science Educator 25(1):10-21.
Lee J, Shibley J (2015). Trajectories of self-perceived math ability, utility value and interest across middle school as predictors of high school math performance. Educational Psychology 37(4):438-456.
Levine S, Moore DP, Bene E, Smith MW (2023). What if it were otherwise? Teachers use exams from the past to imagine possible futures in the teaching of literature. Reading Research Quarterly 58(1):5-24.
Logan J, Burdick J (2016). School segregation, charter schools, and access to quality education. Journal of Urban Affairs 38(3):323-343.
Lozano O (2022). Analysis of the supply elasticity in basic formal education. Human Review 13(1):2-18.
Marlatt R (2018). Get in the game: Promoting justice through a digitized literature study. Multicultural Perspectives 20(4):222-228.
Martinez W, Esquivel I (2017). Reading comprehension strategies of English for high school students, supported. Virtual Campus Magazine 6(1):9-22.
McCulloch A, Allen K, Kenney R (2013). What to trust: Reconciling mathematical work done by hand with conflicting graphing calculator solutions. School Science and Mathematics 113(4):201-210.
Meeter M (2022). Predicting retention in higher education from high-stakes exams or school GPA. Educational Assessment 28(1):1-10.
Merke F, Stuenkel O, Feldmann A (2021). Reimagining regional governance in Latin America. Carnegie Endowment for International Peace. Available at: https://carnegie-production-assets.s3.amazonaws.com/static/files/Merke_Stuenkel_and_Feldman_Latin_America.pdf
Mullis IV, Martin MO, Kennedy A, Trong K, Sainsbury MP (2015). Assessment frameworks. TIMMS and Pirls International Study Center, Boston College.
National Ministry of Education (2009). Student dropout in Colombian higher education: Monitoring methodology, diagnosis and elements for prevention. Bogotá: National Printing Office of Colombia.
Organisation for Economic Co-operation and Development (OECD) (2019). The OECD's work on education and skills. Directorate of Education and Skills. Available at: https://www.oecd.org/education/El-trabajo-de-la-ocde-sobre-educacion-y-competencias.pdf
Onosko J (2011). Race to the top leaves children and future citizens behind: The devastating effects of centralization, standardization, and high-stakes accountability. Democracy and Education 19(2):1-11.
Patton C, Cherng H-Y (2018). Accountability reform and responsive assessment for immigrant youth. Theory Into Practice 57(2):119-126.
Peñaranda J, Aragón A, Micolta G (2014). Description and preliminary evaluation of a program for improving chemistry. Journal of Chemical Education, 91:1439-1445.
Pollio M, Hochbein C (2015). The association between standards-based grading and standardized test scores as an element of a high school reform model. Teachers College Record 117(11):1-28.
PRISMA (2020). PRISMA 2020 flow diagram. Available at: http://www.prisma-statement.org/documents/PRISMA_2020_flow_diagram_new_SRs_v1.docx
Programme for International Student Assessment (PISA) (2006). Technical report. Paris: OECD. Available at: https://www.oecd.org/pisa/data/42025182.pdf
Purnomo H, Sa'dijah C, Hidayanto E (2022). Development of instrument numeracy skills test of minimum competency assessment (MCA) in Indonesia. International Journal of Instruction 15(3):635-648.
Ramadhana A, Susetyo B, Indahwati (2021). Development of science items for minimum competency assessment in Indonesia. International Journal on Advanced Science, Engineering and Information Technology 11(2):501-507.
Ramírez R, Carcausto W (2024). Factors associated with student dropout in Latin American universities: Scoping review. Journal of Educational and Social Research 14(2).
Ramos C, Cóndor O, Cruz J (2023). Evaluation of online learning platforms in Latin America. Emerging Science Journal 6:253-263.
Ross L, Chapman KP, Dorn S, Casanova CR (2023). Opting out of standardised tests at the secondary level—A geographic analysis of Colorado. Educational Researcher 9(1):1-21.
Sánchez M, Pedreño M, Ponce A, Navarro F (2023). And, at first, it was the research question… The PICO, PECO, SPIDER and FINER formats. Espiral. Cuadernos del Profesorado 16(32):126-136.
Sharif A, Karimianpour G, Tagavi H, Amini M, Alafasghari F (2022). Psychometric properties of school climate scale among secondary school students. Journal of School Psychology 10(4).
Smith M (2018). New multiple-choice measures of historical thinking: An investigation of cognitive validity. Theory & Research in Social Education 46(1):1-34.
Smith Z, Flax M, Becker S, Langberg J (2023). Academic motivation decreases across adolescence for youth with and without attention-deficit/hyperactivity disorder: Effects of motivation on academic success. Journal of Child Psychology and Psychiatry 64(9):1303-1313.
Southern Adventist University McKee Library (2023). Google & Google Scholar: Boolean operators. https://southern.libguides.com/google/boolean.
TIMSS & PIRLS (2009). TIMSS Advanced 2008 technical report. Available at: https://timssandpirls.bc.edu/timss_advanced/tr.html
United Nations Educational, Scientific and Cultural Organization (UNESCO) (2005). UNESCO's conceptualization of quality: A framework for understanding, monitoring and improving educational quality. http://www.unesco.org/education/gmrdownload/chapter1.pdf
United Nations Educational, Scientific and Cultural Organization (UNESCO) (2020). Sustainable Development Goals. Goal 4: Ensure inclusive, equitable, quality education and promote lifelong learning opportunities for all. https://www.un.org/sustainabledevelopment/es/education/
UNICEF (2020). Education and learning – Educational quality: Girls, boys and adolescents have the right to go to school andlearn. Available at: https://www.unicef.org/colombia/retornoalaescuela#:~:text=%E2%80%9CIr%20a%20la%20escuela%20y,los%20ni%C3%B1os%2C%20ni%C3%B1as%20y%20adolescentes
Uretsky M, Henneberger A (2023). Expanding the understanding of high school non-graduates through a comparison of high school dropouts and persisters. Journal of Education for Students Placed at Risk (JESPAR) 28(2):236-257.
Vinajera-Zamora A, Gaus N, Rodríguez-Martínez Y (2023). Framework and analysis of key performance indicators in Cuban higher education. Journal of Hispanic Higher Education 22(2):205-218.
Wiyarsi A, Rizanatul A, Supriadi D, Damanhuri MI (2019). A test of analytical thinking and chemical representation ability. Cakrawala

Pendidikan 38(2):228-242.
World Bank (2008). The quality of education in Colombia: An analysis and some options for a policy program. Washington, D.C.: World Bank. https://documents1.worldbank.org/curated/en/914251468018051264/pdf/5435101gray0co1c0dislosed0101221101.pdf
World Bank (2010). Colombia: Quality of education in Colombia. Washington, D.C.: World Bank. https://documents1.worldbank.org/curated/en/914251468018051264/pdf/5435101gray0co1c0dislosed0101221101.pdf
Yao D, Lin J (2023). Identifying key factors influencing teaching quality: A computational pedagogy approach. Systems 11(9):455-479.