

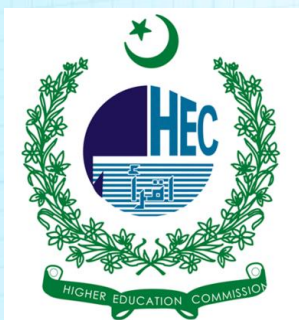
Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

<https://llrjournal.com/index.php/11>

**International Testing Agencies: A Comparative Analysis of
Global Assessment Practices**



***¹Zainab Qamar**

²Noor-ul-Ain Abbas

³Sohail Ahmad

^{*1}Research Associate, Faculty of Social Sciences and Humanities, Minhaj University Lahore, Pakistan & PhD Scholar, Institute of Education and Research, University of the Punjab, Lahore, Pakistan.

zainab.ra@mul.edu.pk

²PhD Scholar, Department of Secondary Education, Lahore College for Women University, Lahore, Pakistan.

dr.noorulainabbas440@gmail.com

³PhD Scholar, Institute of Education and Research, University of the Punjab, Lahore, Pakistan.

sohailahmad606@gmail.com

Abstract

This study explores how international testing agencies and large-scale testing influence accountability, educational quality, and student adaptability. This study compares the performance of agencies such as Educational Testing Service (ETS), Universities and Colleges Admissions Service (UCAS), the National Testing Agency of India (NTA), and the China Gaokao system to worldwide benchmark tests such as PISA, TIMSS, and PIRLS. A qualitative comparative study design was applied, including secondary sources (institutional reports, peer-reviewed publications, and global evaluation databases). Thematic data analysis revealed parallels and contrasts in governance structures, technology integration, psychometric consistency, transparency, and influence on international policy. The discoveries show that whereas the agencies such as ETS, UCAS place a premium on psychometric validity, and centralized efficiency, others such as NTA, Gaokao call on the reputation of large-scale national testing, but remain confronted by evenhandedness and impartiality. International tests such as PISA, TIMSS and PIRLS offer international standards but may be biased and excessively standardized. The paper concludes that international testing agencies play a vital role in education reform in the world, but should strike a balance between standardization and inclusiveness to achieve fair results.

Keywords: Testing, International Testing Agencies, ETS, UCAS, NTA, Gaokao, PISA, TIMSS, PIRLS, Educational Assessment, Comparative Analysis

Introduction

The evaluation has become one of the foundations of the current education system, as it has become the primary tool of determining the student learning process, developing the curriculum, and holding everyone accountable (Black & Wiliam, 2018). When it comes to globalization, assessment has to operate beyond the boundaries of nations since it helps in the comparability of various education systems. Through standardized testing, international testing agencies and large-scale testing allow countries to track the quality of education, its strong points and flaws, and to ensure that their policies are aligned with those of other countries. Therefore, the role of assessment has not only been a classroom instrument, but now comes as a

worldwide educational reformation power.

A few leading agencies emphasize on diversity in international testing practices. In the United States, educational testing service (ETS) develops world-famous tests such as TOEFL and GRE, and they focus on psychometric rigor (Phelps, 2016). Universities and Colleges Admissions Service (UCAS) in the United Kingdom is a single admissions centre that makes it easier to enter higher education in the country, both as a domestic and international student (UCAS, 2023). In addition to agency-based testing, national testing agencies like the National Testing Agency in India and Gaokao in China are also high-stakes tests with massive social and educational impact (Gupta, 2020; Zhao, 2014). They are accompanied by cross-country comparisons with PISA, TIMSS, and PIRLS, which can be used to understand the achievement of students and equity within the system (Mullis et al., 2020; OECD, 2019).

These agencies have the great policy relevance. The international tests affect governments in terms of curriculum reforms, training of teachers as well as equity. As an example, PISA has been used to evaluate the quality of education systems, and such countries as Germany or Japan reform their policies depending on the results (Carnoy, 2015). TIMSS and PIRRS offer data related to specific subject areas that inform national policy in mathematics, science and literacy. Precisely an outcome, worldwide testing bodies not only inspect achievements but also reframe instructional objectives, focusing on competencies and abilities that contribute to competitiveness worldwide.

Despite their contributions, international testing agencies have significant challenges to overcome. Discrimination in the development of tests, bias in having access to study information, and the overwhelming majority of testing preparation organizations constitute numerous matters which require to be taken seriously confronted (Au, 2016). Furthermore, technological inequity hinders the implementation of virtual screenings, especially among nations in the developing world. Demanding evaluations, especially Gaokao assessments as National Testing Agency examinations, place learners to tremendous quantities of anxiety, which may trigger mental anguish (Zhao, 2014). All of these challenges indicate that, whereas worldwide tests enhance uniformity and comparative advantage, they additionally pose with the potential of increasing inequalities and diminishing opportunities for

education.

The following research is valuable because historically there are currently insufficient serious comparative analyses of various international testing agencies. As absence of comprehensive examination, governments might interpret foreign assessment materials in a distorted or inaccurate way, potentially which may result in counterproductive solutions. Lack of such research may lead to excessive focus on rankings, lack of emphasis on the local contexts, and increased separation between the privileged and disadvantaged groups. Thus, the research is timely to give insights into how international testing agencies work, what are their areas of strength and weakness, and how their practices can be enhanced to make them fair, inclusive, and meaningful in global benchmarking.

Objectives of the Study

1. To examine the governance, structure and role of international testing agencies.
2. To compare the agencies like ETS, UCAS, NTA and Gaokao with large scale international assessments like PISA, TIMSS, and PIRRS.
3. To assess the strengths, weaknesses and policy implications of these systems.
4. To recommend ways to foster fairness, inclusiveness and global relevance.

Research Questions

RQ1. What roles do international testing agencies play in ensuring quality, accountability, and mobility in education?

RQ2. How do ETS, UCAS, NTA, and Gaokao differ in governance, structure, and practices?

RQ3. What impact do large-scale assessments such as PISA, TIMSS, and PIRLS have on educational systems worldwide?

RQ4. What challenges limit the effectiveness and fairness of international testing agencies?

Literature Review

International testing agencies and large-scale assessments have been extensively studied over the past decades. Research has focused on their structure, psychometric properties, equity, cultural fairness, policy influence, and the trade-offs involved. This section reviews prior work on ETS, NTA (India), PISA / TIMSS / PIRLS, and other related assessments. It also highlights gaps in literature and under-studied areas.

The Educational Testing Service (ETS) is often treated as a benchmark for rigorous testing practice globally. According to ETS's own documentation, they maintain International Principles for the Fairness of Assessments, guidelines to ensure that test items are appropriate across cultures, and to ensure accessibility (including for test-takers with disabilities) (ETS, n.d.). For example, ETS publishes fairness review publications that show how they examine their tests for bias, translate or adapt items for English-language learners, and ensure that different groups are not disadvantaged (ETS, n.d.). In psychometric literature, ETS tests like TOEFL, GRE and others are studied in terms of validity and reliability. Researchers point out that ETS uses both Classical Test Theory (CTT) and Item Response Theory (IRT) models in test development, calibration, and validation (Kunnan, 2004; Phelps, 2016).

Nevertheless, multiple research studies have shown that even trustworthy agencies, for instance the Examination Testing Service (ETS), may encounter intercultural and demographic discrimination when their tests are administered or executed in surroundings other than the person's residence. To illustrate, objects that were first created in the context of the U.S. or English language, or contain some cultural cues, or the language styles will sometimes incorporate these cultural cues or styles that are familiar to some groups of people than others, which hinges on performance despite some knowledge of the domain (Abedi & Lord, 2001; Hambleton & Patsula, 1998). Such results are indicative of the fact that fairness cannot merely be attributed to a strong psychometrics, but rather to thoughtful cultural adaptation and continued studies on measurement invariance.

The National Testing Agency (NTA) in India is a comparatively new centralized organization (created in 2017), which is meant to streamline and harmonize entrance testing like NEET, JEE Main, UGC-NET. Its strategic plan explicitly states that tests will be “scientifically prepared based on rigorous research ... fair, valid and reliable” and that research staff will include psychometricians and assessment experts (NTA, Strategic Plan). However, practice has shown several controversies and criticisms. A major issue has been the transparency of result declarations, slips in timelines, and accusations of irregularities. For NEET-UG 2024, for example, the NTA awarded “grace marks” to a large number of candidates because of time lost, which helped many achieve perfect scores; this

raised questions about whether criteria for awarding grace marks were clearly established or communicated (Frontline, 2024; Times of India, etc.).

Legal challenges have also been increasing. Since its establishment, over a thousand cases have been filed against NTA in Indian courts, many about paper leaks, alleged miscalculations, or inconsistencies in evaluation or result normalization methods. Reform efforts have been proposed and partially implemented. For example, a high-level committee has been formed to recommend changes to NTA's processes, digital infrastructure, IT security, and transparency practices.

Large international assessments like PISA (conducted by the OECD), TIMSS and PIRLS (by IEA) have been central to cross-national policy comparisons. These serve multiple purposes: evaluating student achievement in subjects like math, science, reading; diagnosing system-level strengths and weaknesses; encouraging reforms; and informing stakeholders (teachers, policy makers, public) about performance vis-à-vis international peers. An important line of literature has focused on measurement invariance and item bias. For example, a study by Alatlı (2021) examined cross-cultural measurement invariance in the science literacy domain of PISA 2015 among Australia, France, Singapore, and Turkey. They found that while the overall structure held (i.e. the test measured roughly the same constructs across countries), a large proportion of items displayed differential item functioning (DIF), meaning that some items favour certain populations (Australia-Singapore, Australia-France comparisons showed ~35% items with DIF). Similarly, studies on the reading skills items in PISA 2018 (Australia, China-B-S-J-Z, Turkey) revealed that certain items did not maintain invariance across countries, and bias was observable in translation and content familiarity (Balıkesir University etc.). Eryılmaz, Rivera-Gutiérrez & Sandoval-Hernández (2022) used a theoretical approach based on Bourdieu's cultural reproduction theory to create new socio-economic scales (economic, cultural, social capital) and tested measurement invariance. They found that the socio-economic background scale often does not map equivalently across diverse contexts, which makes comparisons problematic if not adjusted or accounted for carefully.

Also, the literature addressed the variances between category proficiency regarding the PISA test. A research investigation entitled "Understanding the Disparities in Mathematics Performance: An Interpretability-Based Examination"

(2025) applied PISA information along with techniques such as the SHAP technique (Shapley Additive Explanations) to evaluate comprehension easy access, intellectual curiosity, ethnicity, along with place of residence to be important explanations regarding disparity regarding pupil accomplishment in mathematics.

Comparative analyses of foreign testing agencies and large-scale trials demonstrate plenty of fundamental characteristics. The majority of the concerns revolves around the mismatch between the framework of governance and implementation. Agencies which include ETS and the NTA within India seem extremely detailed throughout their long-term objectives with regard to how they can appear unbiased, precise and authentic, however what actually occurs on the field indicates an extensive number of inequalities. As an example, although NTA has assured to enhance psychometrics practices through the hiring of experts, its activities remain plagued with problems like incoherence, delay and claims of anomaly, which portrays a lack of policy-practice fit. Psychometric rigor and significance of measurement invariance is another major theme. Intel international tests such as PISA, TIMSS, and PIRLS are based on cross-country comparisons, however, studies have continuously cautioned that without invariance, test items can be biased towards certain cultural or national groups. Articles like Alatlí (2021) show that the occurrence of differential item functioning (DIF) is still rampant and thus, challenges the fairness and validity of such global standards. It also brings about another factor as transparency and public trust that is important in keeping the credibility. The studies and reports highlight that to maintain legitimacy of high-stakes testing, a series of processes like release of answer keys, reasoning of the need to compensate lost examination time, and transparent management of unfair means are vital. Accessibility is also a crucial dimension as well as equity. Results of PISA and TIMSS indicate that socio-economic status and urban-suburban factors and access to resources such as technology and books continue to be strong predictors of student achievement. Standardized tests are likely to perpetuate rather than alleviate existing differences in the absence of specific accommodations for such variances. When assessments are translated, such as Turkey's input in PISA, items tend to behave differently than in their original presentations, generating validity difficulties. Similarly, PIRLS revealed that reading comprehension is inextricably linked to

language and cultural context, making cross-country comparisons difficult. Finally, the study highlights the role of international evaluations in policy, curriculum, and instructional approaches. PISA, in particular, has worked as a reformatory catalyst, forcing countries to revise curricula and teacher training, as well as implement new assessment methods.

Despite extensive research on international testing agencies and large-scale assessments, several gaps remain under-explored. For instance, examinations like China's Gaokao have profound educational and social implications, yet empirical studies in English remain limited, particularly regarding issues of fairness, test-related stress, curriculum narrowing, and item bias. Another gap lies in the scarcity of longitudinal research; most studies rely on cross-sectional data, leaving unanswered questions about how international assessment scores relate to long-term outcomes such as university achievement, employment opportunities, and broader life trajectories. The rapid shift toward digital assessment has also outpaced scholarly inquiry, with relatively little evidence on how online versus paper-based formats influence fairness, performance, or perpetuate digital inequalities. The adaptation of assessments to regional and local contexts, especially in non-English-speaking or resource-constrained settings, remains an area requiring further case-based research to ensure cultural and linguistic fairness. Collectively, these gaps highlight the need for more nuanced, diverse, and context-sensitive research to strengthen the validity and inclusivity of international testing practices.

Research Design and Methodology

The present study adopted a qualitative comparative research design, as the primary aim was to explore and contrast the structures, functions, and challenges of international testing agencies and large-scale assessments. The research was exploratory in nature, seeking to generate insights rather than test hypotheses, and therefore relied on secondary data sources. The data was collected through the official websites, quarterly surveys, parliamentary seminars, journals with peer-review, as well as worldwide scoring forums such as the OECD and International Energy Agency. A broad spectrum of data sources guaranteed that every bit of the details that were provided remained exhaustive and genuine.

The collected data were then analyzed by using thematic content analysis approach.

Major themes and sub-themes were identified, and coded systematically. Major themes include governance structure, functions and challenges. Patterns and differences were then examined over time by agency (ETS, UCAS, NTA, Gaokao, and high-stakes tests including PISA, TIMSS, and PIRLS). A comparative table was developed to add further clarity which allowing people to review administrative procedures against one another along with observe their differences as they link to the deeper consequences for worldwide educational frameworks. The methodology provides the ability to understand the strengths and limitations of international testing frameworks fine-tuning.

Results and Discussion

The first theme in the comparative analysis has to do with how international testing agencies are set up and operate. Such organizations as the Universities and Colleges Admissions Service (UCAS) in the United Kingdom and the Educational Testing Service (ETS) in the United States are extremely specialized and centralized, in contrast to the fragmented systems seen in the majority of developing countries. ETS has established a standardized and consistent testing system that creates and administers internationally recognized exams, such as the Praxis, GRE, and TOEFL (ETS, 2020). established in India with the intention of centralizing professional examinations, which were previously thought to be anomalous, in order to increase responsibility and decrease anomalies (Gupta, 2020).

The second theme is associated with governance and transparency. International testing agencies spend a lot of money to ensure high standards of accountability and confidence among the people. To ensure reduced malpractices, ETS and UCAS, among others, have robust security measures, psychometric validation processes and review committees (Kuncel & Hezlett, 2010). Digital monitoring, biometric verification is another more recent change that NTA implemented to prevent corruption and enhance trust in high-stakes exams (Gupta, 2020). With PISA, TIMSS, and PIRRS, the same applies, with OECD and IEA supervising the tests and releasing technical reports, frameworks, and their item-release samples to guarantee transparency and credibility based on research (Mullis et al., 2020; OECD, 2019). Unlike the systems where accountability is poor, such international agencies enhance their legitimacy by creating trust through strict

governance systems and communication of the assessment process.

The third theme is that of psychometric rigor and test validity. Assessments like PISA, TIMSS, and PIRLS are also international with advanced psychometric frameworks like Item Response Theory (IRT) used to calibrate items and provide fairness to the different cultural and linguistic background (Embretson & Reise, 2000). ETS has also continually done research on psychometrics to narrow item pools, study different item functioning, and test predictive reliability of ETS assessments (Phelps, 2016). Although the issues of cultural bias still exist, the current research and adaptations are designed to reduce the levels of inequities and preserve the test validity. By comparison, systems that do not contain such strict structures do not have an easy time with the notions of fairness, reliability, and international acceptance. In this way, psychometric validity is another common strength of foreign testing organizations.

The fourth theme is integration and innovation of technology. The international agencies have been the first ones to utilize digital testing platforms so as to enhance accessibility and cut logistic hurdles. Decades ago, ETS has switched several of its most popular academic examinations, such as GRE and TOEFL, to computer-based testing, and is still testing AI-based grading and adaptable testing platforms (Zhao, 2018). In India, the NTA has introduced the use of computer-based exams at JEE and NEET level in large scale that are facilitated by secure digital systems and real time monitoring. UCAS takes advantage of online portals to facilitate the university application process and online platforms are used by organizations such as OECD and IEA to administer PISA and TIMSS in various countries. These inventions reflect the ability of technology to increase efficiency, transparency and scale of international testing processes.

The last theme is concerned with the policy impact and reform potential of international testing agencies. Governments often rely on the results of PISA, TIMSS, and PIRRS as a measure of national education policy and begin using them to implement changes in curriculum, pedagogy, and teacher training (Carnoy, 2015). As an illustration, in accordance with the TIMSS reports, science and mathematical modifications to the curriculum are being implemented in several other states, whereas the findings from PISA have influenced policy debates regarding equitable educational opportunities and standards. For an identical

point, UCAS and ETS having an immediate influence with the construction of admittance as well as testing processes that shape educational degrees throughout all over the globe. These entities demonstrate that uniform testing cannot exclusively utilized as quantify, but additionally for achieving modifications to the framework. Still, researchers point out because depending extensively on international norms could harm developing curricula and promote standings higher than indigenous goals for instruction.

The comparative research highlights the capabilities of international testing organizations with regard to that administrative accountability, cognitive accuracy, innovation in technology, and governmental usefulness. Nevertheless, fears including as stereotypes, nervousness about tests, and the technological gap indicate the capabilities of even familiar organizations remain limited. In conclusion, international assessment frameworks highlight the way successful management facilitates the advancement of robust assessment techniques. Regional benchmarking may help foster esteem and legitimacy, which at first has significance throughout nations which require to improve their instructional consideration implements.

Tabel 1: *Comparative Insights of the Study*

Theme	International Agencies (ETS, UCAS, NTA, PISA/TIMSS/PIRLS)	Key Insights
Governance & Structure	Centralized and specialized ETS (USA) administers examinations such as GRE/TOEFL; unified OECD/IEA PISA/TIMSS Global Frameworks; UCAS (UK) is a uniting body in admissions; NTA (India) administers multiple large-scale examinations.	Centralized structures promote efficiency, standardization, and transparency, unlike fragmented national systems.
Governance & Transparency	ETS/UCAS has high supervision and responsibility; NTA has digital surveillance and grievance-redressal; OECD keeps processes of PISA/TIMSS reviewed externally.	Strong governance mechanisms build public trust and credibility in results.

Psychometric Rigor	ETS uses Item Response Theory (IRT), DIF analysis, and continuous validation; PISA/TIMSS stresses measurement invariance across nations so as to be fair; NTA is progressively making psychometric improvements.	Advanced psychometric practices ensure fairness, reduce bias, and enhance predictive validity.
Technological Integration	Computer based testing was pioneered by ETS, NEET/JEE is administered online by NTA, admissions were digitalized by UCAS, and PISA lately administrations were majorly conducted using digital, adaptive forms.	Technology enhances efficiency and security, but raises issues of equity in resource-constrained contexts.
Policy Impact & Reform Potential	Curriculum and policy changes in different countries around the world are based on PISA results; admission and hiring criteria are changed by UCAS and ETS, and accountability and modernization are the priorities of NTA reforms.	Large-scale assessments shape educational policies, but excessive reliance may lead to “teaching to the test.”

Table 1 summarizes the thematic findings through an in-depth comparison of worldwide testing organizations, underlining whether transparency, accountability, cognitive reliability, the integration of technology, as well as governmental consequences on universal methods for assessment. Research illustrates effective hierarchical schemes, psychometric refinement, and technological advancement boost trustworthiness along with governmental effectiveness, though uniformity and disparities in inequality represent major concerns.

Challenges

Despite the widespread adoption of international assessments, testing agencies face significant challenges that compromise fairness and inclusivity. One major issue is measurement invariance: studies of PISA’s science literacy test (e.g. Alatl, 2020) and PISA 2018 reading skills (Alatl, 2022) reveal that many items show Differential Item Functioning (DIF) across countries, especially between those using translated versions of tests, undermining comparability. Socio-economic and regional disparities also

persist; for example, in China's Gaokao system, regional inequalities intensify the psychological burden on examinees, with students from less privileged areas experiencing more stress and fewer coping resources (Chan, Guo, Lu, & Zhang, 2025). The psychological effects are compounded by academic pressure: a recent study found that stress negatively correlates with Gaokao student academic performance, with academic self-efficacy and positive psychological capital mediating this relationship (Lu, 2025). Another key challenge is the cost burden of high-stakes tests like IELTS, where students incur repeated fees, travel expenses, and potential financial loss due to retakes, especially when aiming for minimum module requirements (IELTS cost study, Saudi Arabia; 2018). Additionally, as assessments increasingly shift to digital platforms, there is growing concern over access in resource-poor settings and about digital divides that may exacerbate existing inequities (studies on PISA 2022 online information questionnaire invariance; see Education and Information Technologies, 2024). All of these results demonstrate the fact that in addition to provide global relevance as well as impartiality, most reputable international assessment systems need to address concerns regarding discrimination, gender equity, psychological wellness, and financial barriers.

Conclusion

The research revealed major global testing agencies including the ETS, UCAS, and NTA, in addition to international tests such as PISA, TIMSS, and PIRLS, are establishing them as legitimate, transparent, and psychometrically secure models of large-scale educational evaluation. The tests are efficient and fair because they're structured, founded on prevailing testing theories such as Item Response Theory (IRT), and are connected using technological portals; their findings can occasionally influence national and international policies regarding education. Despite these positive qualities, equity of access, cultural bias, a heavy emphasis overall controlled evaluations, and the demands that these agencies impose on the educational system constitute numerous drawbacks. Overall, this contrast emphasizes the relevance regarding implementing the most effective techniques established by different worldwide precedents as keeping aware of specific prerequisites in order to ensure assessment procedures help promote equality, creative thinking, and ongoing education rather than simply testing knowledge.

Recommendations

1. National testing bodies should use modern methods like Item Response Theory (IRT), checking for unfair questions (DIF analysis), and regularly reviewing their tests to make sure they are fair and accurately predict student performance.
2. Testing agencies need to set up transparent systems and allow independent audits. This will build public trust and show that the test results are reliable.
3. Exams should not only focus on memorization. They must be aligned with broader educational goals so that assessments support real learning, not just rote practice.
4. Stronger international collaboration with organizations like ETS, UCAS, and OECD is needed. This way, national testing systems can stay updated and closer to global standards.
5. Continuous research is important to make tests more inclusive, reduce cultural bias, and focus on essential 21st-century skills such as problem-solving, creativity, and critical thinking.

References

- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14(3), 219–234.
https://doi.org/10.1207/S15324818AME1403_2
- Alathi, B. (2021). Cross-cultural measurement invariance of PISA 2015 science literacy assessment: Evidence from Australia, France, Singapore, and Turkey. *International Journal of Assessment Tools in Education*, 8(2), 227–247.
<https://doi.org/10.21449/ijate.835969>
- Au, W. (2016). *Meritocracy 2.0: High-stakes, standardized testing, and the education of the new working class*. Routledge.
- Balıkesir University. (2022). Cross-cultural validity of PISA 2018 reading literacy items: A Turkish case study. *Journal of Measurement and Evaluation in Education and Psychology*, 13(1), 45–62.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575.
- Braun, H., & Kanjee, A. (2006). Using assessment to improve education in developing nations. *Educational Measurement: Issues and Practice*, 25(4),

31–46.

- Carnoy, M. (2015). International test score comparisons and educational policy: A review of PISA's impact on national education systems. *Teachers College Record*, 117(1), 1–26.
- Chan, S., Guo, Y., Lu, X., & Zhang, L. (2025). Regional inequality, exam stress, and coping among Gaokao students in China. *Asia Pacific Education Review*, 26(2), 145–160. <https://doi.org/10.1007/s12564-024-09789-4>
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum.
- Erylmaz, A., Rivera-Gutiérrez, M., & Sandoval-Hernández, A. (2022). Measuring socio-economic status in cross-national surveys: A Bourdieu-inspired approach using PISA data. *Large-Scale Assessments in Education*, 10(6), 1–18. <https://doi.org/10.1186/s40536-022-00140-8>
- ETS. (2020). *About ETS*. Educational Testing Service. Retrieved from <https://www.ets.org>
- ETS. (n.d.). *ETS international principles for fairness review of assessments*. Educational Testing Service. Retrieved from <https://www.ets.org>
- Frontline. (2024, June). NTA and the NEET controversy. *Frontline: India's National Magazine*. Retrieved from <https://frontline.thehindu.com>
- Gupta, A. (2020). Standardizing admissions: The role of India's National Testing Agency. *Economic and Political Weekly*, 55(14), 25–30.
- Gupta, R. (2020). National Testing Agency and reforms in Indian higher education entrance examinations. *Journal of Educational Studies*, 18(2), 45–59.
- Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, 45(1–3), 153–171. <https://doi.org/10.1023/A:1006941729637>
- IELTS Cost Study. (2018). Financial implications of repeated IELTS testing among Saudi students. *International Journal of English Language Testing*, 6(2), 55–67.
- Kuncel, N. R., & Hezlett, S. A. (2010). Standardized tests predict graduate students' success. *Science*, 327(5968), 1081–1082. <https://doi.org/10.1126/science.1186146>

- Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), *European language testing in a global context* (pp. 27–48). Cambridge University Press.
- Lu, X. (2025). Academic stress and performance in China's Gaokao: The mediating roles of self-efficacy and psychological capital. *Journal of Adolescence*, 100, 45–57. <https://doi.org/10.1016/j.adolescence.2024.12.004>
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. International Association for the Evaluation of Educational Achievement (IEA).
- OECD. (2019). *PISA 2018 assessment and analytical framework*. OECD Publishing. <https://doi.org/10.1787/b25efab8-en>
- Phelps, R. P. (2016). *Correcting fallacies about educational and psychological testing*. American Psychological Association. <https://doi.org/10.1037/14954-000>
- Times of India. (2024, June 7). NEET results controversy: NTA faces backlash. *The Times of India*. Retrieved from <https://timesofindia.indiatimes.com>
- UCAS. (2021). *About UCAS*. Universities and Colleges Admissions Service. Retrieved from <https://www.ucas.com>
- UCAS. (2023). *UCAS annual report and accounts 2022*. Universities and Colleges Admissions Service.
- Zhao, Y. (2014). *Who's afraid of the big bad dragon? Why China has the best (and worst) education system in the world*. Jossey-Bass.
- Zhao, Y. (2018). *Reach for greatness: Personalizable education for all children*. Corwin Press.