

Equity, Access, and Barriers in English Language Proficiency Testing: A Systematic Review of Bangladeshi Students' Pathways to International University Admissions and Scholarships

Sharmin Rahman Bipasha ^{1*} and Md Nahidul Islam ²

¹ Department of English, CCN University of Science & Technology, Cumilla, Bangladesh.

² Department of Public Administration, Comilla University, Cumilla, Bangladesh.

ABSTRACT

English language proficiency tests (ELPTs) like IELTS, TOEFL, PTE and the Duolingo English Test (DET) serve as the official gateway to overseas admission and scholarships for Bangladeshi students, but no synthesis has been done to understand whether the gateway is fair or not. This qualitative descriptive systematic review (according to PRISMA 2020 and ENTREQ) synthesized 38 studies from 2011 to 2026 that were found using Scopus, Web of Science, ERIC, PsycINFO, Google Scholar and citation searching. Thematic synthesis (Braun & Clarke, 2006) was used to reveal three ways in which ELPTs perpetuate inequality, based on critical language testing theory (Shohamy, 2001; McNamara & Ryan, 2011) and theories of educational stratification. First, there are differential costs of access based on class and rural–urban lines due to direct and indirect costs (test fees USD 205–250/sitting, coaching, resits and travel). Second, test preparation is differentiated: coaching improves IELTS scores by about half a band, but not the underlying level of proficiency, and bought test preparation translates into acceptable credentials for families. Third, cut score systems are still in use even though they have low predictive validity (meta-analytic $r = .231$). The review explains that the tests serve two purposes: they measure communicative competence; and the conditions for demonstrating competence are socioeconomically distributed, and scores thus measure competence and position. Actionable implications are provided for the Bangladesh Ministry of Education, universities and scholarship providers, and test providers. A research agenda is suggested, which looks forward.

ARTICLE HISTORY

Received 11 August 2026

Revised 15 October 2026

Accepted 16 October 2026

Online First 16 October 2026

KEYWORDS

Bangladesh; Critical language testing; English language proficiency testing; International scholarships; Systematic review

1. Introduction

1.1 Background

Proficiency in English on a standardized test has become a crucial stumbling block in the way of students from the Global South who want to pursue higher education and scholarships abroad (Hamid et al., 2019b). English language proficiency testing (ELPT) is part of the formal gatekeeping system of universities in the United Kingdom, the United States, Australia, and Canada, as well as of multinational scholarship programmes (Deygers et al., 2018). In many cases, Bangladeshi applicants' admission and eligibility for the funding, without which international study is impossible, is determined by a single band score. While a global test-preparation industry and digital alternatives have arisen, the legacy tests are still the dominant ones (Trenkic & Hu, 2021). Medium of instruction (MOI) certificates, DET scores, or in-house assessments are now accepted by some institutions, but IELTS and TOEFL are still the most popular evidence (Yao, 2023). This persistence can be attributed to institutional inertia and the commercial interests of a testing industry that makes billions of dollars every year in testing fees, materials, and coaching (Hamid et al., 2019b). Importantly, validity statements for these tests are typically made at the instrument level, but the implications of the tests are at the level of use – who can sit, re-sit, prepare; who will have scores recognized (Kane, 2013; Pill et al., 2025).

1.2 The Bangladeshi Context

English has an ambiguous status in Bangladesh: it is mandatory in all schools, but the quality, legitimacy, and access to English are very uneven (Hamid & Baldauf, 2011; Hamid & Erling, 2016). There are significant disparities in teacher qualifications, materials, and exposure to English between urban English-medium schools or well-resourced schools and rural, under-resourced public schools (M. M. Rahman et al., 2019), and even students entering private universities have weak spoken English (Islam & Stapa, 2021). The availability of private tutoring, which is proven to boost test scores, is related to family income, parental education, and where they live (Ali et al., 2020), creating a de facto tiered system of English development, where international admission is less dependent on effort and ability, and more on socioeconomic status and geographic luck.

Such inequalities are replicated in assessment itself, which is high-stakes. Prior studies reveal a “ritualization” of English testing in Bangladesh where compliance and predictability are valued over validity and fairness (Ali et al., 2020), and the culture of teaching to the test from Higher Secondary Certificate to IELTS and TOEFL preparation (Ali & Hamid, 2020; Kh. A. Rahman et al., 2021). The consequences are very high – ELPT results determine admission and scholarships and thus the financial viability of international education.

1.3 Positioning Bangladesh within Global South Debates and Critical Language Testing

Bangladesh is not considered here as a stand-alone example, but as a telling example in the broader Global South debates. Critical language testing views tests as tools of power, and thus their uses, not their items, have social consequences, and validity is not only related to the nature of the test items but also to the justice of the test-based decisions (McNamara & Ryan, 2011; Shohamy, 2001, 2022). From this point of view, the question is not just whether IELTS or TOEFL is a reliable measure of English but whether the use of cut scores as a criterion for mobility is an equitable distribution of opportunity for groups who are not equally positioned in terms of opportunities to prepare for the test (Hamid, 2016; Hamid et al., 2019b). There is evidence to suggest that it does not: the introduction of an English listening requirement in China's national entrance examination increased the rural-urban gap in access to college by 30% (Li et al., 2024); test takers in the world have doubts about the fairness of the IELTS test, which they believe is profit-oriented (Hamid et al., 2019a; Pearson, 2019); and English proficiency is correlated most closely with socioeconomic status in former British and American jurisdictions and poorer economies (Terasawa, 2025). Bangladesh has all these factors: a colonial linguistic legacy, a high level of English divide, high reliance on outbound scholarships, and admissions criteria that are entirely foreign. The Bangladeshi case is thus an analytically sharp test

of the critical thesis, and the results of the analysis can be applied to other South Asian and Global South scenarios that are confronted with the same gate.

1.4 Conceptualizing Equity, Access and Barriers

The operationalization of three constructs in the review is presented in Table 1. Equity is about substantive fairness, not formal equality: equal opportunity to be prepared, equal measurement of underlying proficiency, and outcomes that are not systematically affected by class, geography or gender (Deygers et al., 2018). Access includes material access (fees, test centres, time, transport) and informational/institutional access (awareness of options, quality of guidance and recognition of alternative evidence). The barriers are financial, geographic, linguistic, policy, informational and cultural limitations that limit students' ability to prepare for, take and succeed on ELPTs. The constructs are at the individual level (socioeconomic status, previous schooling, information), the meso level (schools, coaching centres, test centres, scholarship programmes) and the macro level (national policy, test-industry pricing, requirements of destination institutions).

Table 1

Working definitions and analytical dimensions of equity, access and barriers

Concept	Operational definition	Level(s)	Illustrative indicators
Equity	Substantive fairness in the distribution of opportunities and outcomes; fair access to the preconditions of preparation; valid measurement of proficiency; outcomes free of systematic bias by class, geography or gender	Individual; meso; macro	Subgroup score gaps; predictive bias; demographic parity in test access and outcomes; school-quality parity
Access	Material, informational and institutional access to tests, preparation and pathways; availability of dates, locations and alternatives; awareness and guidance	Individual; meso; macro	Affordability; proximity to centres; coaching availability; awareness of options; institutional recognition of alternative evidence
Barriers	Financial, geographic, linguistic, policy, informational and cultural constraints on preparing for, taking and succeeding on ELPTs	Individual; meso; macro	Fees; distance; instruction quality; coaching scarcity; non-recognition of MOI evidence; rigid cut-score policy

1.5 Aim, Research Questions and Scope

The review draws on qualitative, quantitative and mixed-methods evidence on equity, access and barriers in ELPT from the perspective of Bangladeshi students wishing to gain international admission and scholarships, and comparator evidence from other Global South and international contexts where mechanisms are transferable. It has three questions:

RQ1. What are the experiences of equity and inequity in ELPT for international admission and scholarships of Bangladeshi students?

RQ2. What are the individual, institutional (school, coaching centre, test centre), and policy (national, test-provider, destination-institution) level barriers and facilitators to access to ELPT and onward pathways?

RQ3. What do the included studies show with regard to the validity and fairness of applying fixed ELPT cut scores for admission and scholarships, and what are the implications for policy and pedagogy in Bangladesh and similar contexts?

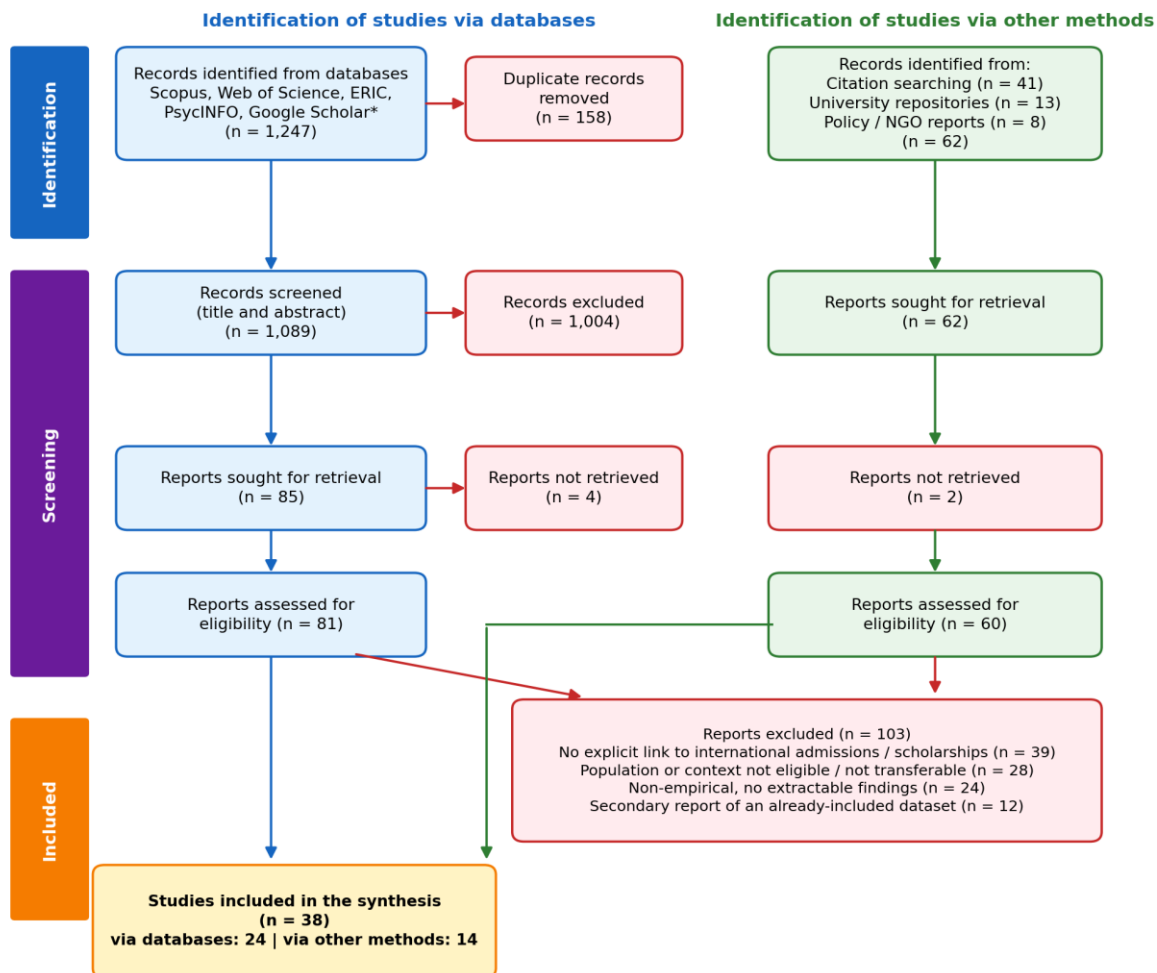
2. Methodology

2.1 Design, Protocol and Reporting Standards

A qualitative descriptive (QD) systematic review design was used. The language of the primary sources is used in a low-inference “straight description” of the phenomena, which is appropriate for questions of the type “what is happening here?” and enables evidence from diverse studies (qualitative studies, surveys, policy analyses) to be synthesized without a strong theoretical over-interpretation (Maeda et al., 2022; Sandelowski, 2000). QD was operationalized at the review level using thematisation (Thomas & Harden, 2008). The review was carried out and reported based on the PRISMA 2020 statement (Page et al., 2021) and ENTREQ framework for qualitative syntheses (Tong et al., 2012). The protocol was designed using the SPIDER tool (Sample, Phenomenon of Interest, Design, Evaluation, Research type), which is more appropriate than PICO for qualitative and mixed-methods evidence (Cooke et al., 2012). No attempt was made to register the PROSPERO as the register only covers reviews that have health-related outcomes; in line with the suggestions made for educational syntheses, the protocol was finalized before the search was conducted, archived with the corresponding author, and is available on request. The selection process is summarized in Figure 1.

Figure 1

PRISMA 2020 flow diagram of study identification, screening and inclusion (adapted from Page et al., 2021)



*Google Scholar: first 100 results per search string (see Section 2.3). Flow diagram follows Page et al. (2021).

2.2 Eligibility Criteria

Eligibility criteria were designed using SPIDER (Table 2). This included Bangladeshi students at secondary or tertiary/graduate level, as studies of other international or Global South cohorts were only included if they investigated the same tests, policies or mechanisms (e.g., predictive validity, coaching effects, cut-score policy) and the authors deemed the findings transferable to the Bangladeshi pathway. Given that RQ3 is about the validity of cut-score use, quantitative predictive-validity and experimental studies were included, as well as qualitative studies (which in an earlier version of the protocol were the only studies included).

Table 2*Inclusion and exclusion criteria operationalized through the SPIDER framework*

SPIDER domain	Inclusion criteria	Exclusion criteria
Sample	Bangladeshi students (secondary, tertiary, graduate); international or Global South cohorts where mechanisms are transferable	Cohorts with no link to Bangladesh or to the tests and policies Bangladeshi applicants face; teacher-only samples
Phenomenon of interest	Experiences, costs, preparation, policies and validity of ELPT (IELTS, TOEFL, PTE, DET) linked to international admission or scholarships	General ESL pedagogy; test preparation without empirical evidence
Design	Qualitative, quantitative, mixed-methods empirical studies; meta-analyses and systematic reviews; critical policy analyses with extractable findings	Opinion pieces and editorials without evidence or systematic argument
Evaluation	Equity-, access- or barrier-related outcomes on pathways to international higher education or scholarships	Domestic-only outcomes with no admissions or scholarship link
Research type	Peer-reviewed articles; book chapters; reports from established institutes; quality grey literature (2011–2026)	News, blogs, course materials, conference abstracts

2.3 Information Sources and Search Strategy

The searches were conducted between January and March 2024 and repeated in February 2026 to ensure consistency in the publication window (2011–February 2026) and the search dates. The earliest publication date was 2011 to get the first ten years of empirical work on English and socioeconomic disadvantage in Bangladesh (Hamid & Baldauf, 2011). Grey literature was identified in the Bangladeshi university repositories, various donor and NGO reports on student mobility, and policy documents. Google Scholar was limited to the first 100 results per string, as the indexing and ranking of Google is not transparent and may exclude relevant results (Haddaway et al., 2015); therefore, it was used as a supplementary source to the structured databases. Search strings were made up of four concept blocks and Boolean operators (Table 3).

Table 3*Search concepts, terms and filters*

Concept	Search terms	Filters
Geography	Bangladesh OR Bangladeshi OR “South Asia”	Language: English or Bangla
Testing	IELTS OR TOEFL OR PTE OR “Duolingo English Test” OR “high-stakes language test”	Subject: Education/Language; Years: 2011–2026
Context	“international admission” OR “university admission” OR scholarship OR “higher education”	Document type: article, chapter, report
Equity	barrier* OR access OR equity OR fairness OR justice OR exclusion OR gatekeeping	Combined blocks; manual repository search

2.4 Study Selection, Data Extraction and Reliability

Titles and abstracts were screened independently by two reviewers, and subsequently full texts were screened against the operationalized criteria, and any disagreements were discussed. Interrater agreement was above the pre-specified threshold ($\kappa \geq 0.60$), with Cohen's κ at the title and abstract level being 0.74 and 0.81 at the full-text level. Hand-searching of reference lists of included studies was done. A standardized extraction form was used to extract information on bibliographic details and setting, participants, design, test(s) examined, admissions or scholarship focus, equity and access concepts, barriers and facilitators, and key findings and authors' interpretations, and was piloted on five heterogeneous studies. Data extraction and second review of all data records were done by one reviewer. In addition to screening, 20% of the findings were randomly double-coded in NVivo 12, and the agreement on the descriptive codes was 88%, with all differences resolved by consensus, and the codebook amended and an audit trail kept.

2.5 Quality Appraisal and Sensitivity Analysis

The methodological quality was assessed using the CASP Qualitative Checklist, which includes 10 domains: aims, design, recruitment, data collection, reflexivity, ethics, analysis, findings, value and limitations (Long et al., 2020). Quantitative and review studies were appraised using the same ten domains as the CASP cohort, and the systematic-review checklists were similarly mapped to the ten domains. Each domain was rated 2 (clearly met), 1 (partially met), or 0 (not met), and a score of 0-20 was created. A gradation of rigour was applied, with scores of ≥ 14 (at least 70% of criteria clearly met) considered high quality, 10–13 moderate and < 10 low, following Long et al. (2020), who warn that CASP's binary format obscures gradations of rigour. The 70% threshold is based on the study meeting most of the criteria and partially meeting the rest, and is similar to the cut-points used in previous educational thematic syntheses (Maeda et al., 2022). Quality was not used to exclude studies, but a sensitivity analysis was conducted to re-run the synthesis without low-quality studies to determine if any theme was reliant on low-quality studies (results in Section 3.1).

2.6 Data Synthesis

Thematic synthesis was carried out using Thomas and Harden's (2008) three stages: inductive line-by-line coding of the findings extracted; grouping of the codes to form descriptive categories, close to the terminology of the primary studies (low inference); and development of analytical themes, integrating across the categories. The dimensions of equity–access–barrier and the conceptualization of language tests as gatekeeping (Hamid et al., 2019b) were used as a priori themes for guiding the development of themes. Differences by test type, destination, level (individual, institutional, policy), and intersectionality were also considered. The analytical stage was based on critical language testing (Shohamy, 2001; McNamara & Ryan, 2011) and on theories of educational stratification, which see the patterns that emerge as a result of the interplay between the system and the families' economic capital as a way of interpreting, rather than just describing, the results. Data coding and theme management were done using NVivo 12.

3. Results

3.1 Overview of Included Studies and Quality Appraisal

Thirty-eight studies met inclusion criteria and are listed in the full list (Supplementary Table S1) along with design, context, test(s), and quality rating for each study to verify inclusion decisions (Figure 1). Eight studies specifically looked at Bangladesh; the other 30 studies compared the same tests, policies, or mechanisms in other contexts (China, Belgium, Australia, United Kingdom, the United States, cross-national Asia, and cross-national Global South). The corpus was designed to include qualitative empirical studies ($n = 9$), quantitative/experimental studies ($n = 13$), mixed methods studies ($n = 3$), systematic reviews and meta-analyses ($n = 4$), and critical policy or conceptual

analyses (n = 9). The most commonly tested test was IELTS (n = 22), followed by TOEFL (n = 12), DET (n = 8), and PTE (n = 4); some studies tested more than one of these tests. Thematically, the studies covered inequity in preparation and access (n = 15), financial barriers (n = 11), institutional and policy gatekeeping (n = 14), and predictive validity (n = 9), and there was overlap.

24 studies were considered high quality (CASP score ≥ 14), 12 moderate, and 2 were low quality (Table 4). Poor scores were primarily due to low levels of reflexivity and reporting of recruitment and analysis in small descriptive studies in Bangladesh. The sensitivity analysis re-ran the synthesis with the two poor-quality studies excluded, and yielded no change in any of the analytical themes: those studies only provided corroborating detail on EMI disparities and spoken-English weaknesses, which were also supported by high-quality studies. Weak evidence is thus not a prerequisite for the synthesis.

Table 4

Summary of quality appraisal (CASP, 0–20 scale) by study type

Study type	n	High (≥ 14)	Moderate (10–13)	Low (< 10)
Qualitative empirical	9	5	2	2
Quantitative/experimental	13	12	1	0
Mixed methods	3	2	1	0
Reviews and meta-analyses	4	2	2	0
Critical policy / conceptual	9	3	6	0
Total	38	24	12	2

3.2 Equity and Inequity in Test Preparation and Access

There was a significant gap in access to good preparation, depending on socioeconomic, geographic, and institutional factors. Hamid and Baldauf (2011) reported systematic disadvantage for rural Bangladeshi learners, which can be attributed to the lack of qualified teachers, English-medium schools, and coaching, while Hamid and Erling (2016) explained this situation as a result of a policy regime that promises ‘English for everyone’ but lacks the teacher capacity to provide this. Terasawa (2025) found that the socioeconomic gradient of English proficiency is highest in exactly the post-colonial and lower-income contexts, like Bangladesh, in her cross-national analysis of 30 Asian jurisdictions. Like instruction, information about IELTS and TOEFL requirements is also unequally distributed: elite urban students are informed of IELTS and TOEFL requirements 6–12 months ahead of time, while public and rural school students are informed about the deadlines within weeks (Ali et al., 2020; Trenkic & Hu, 2021).

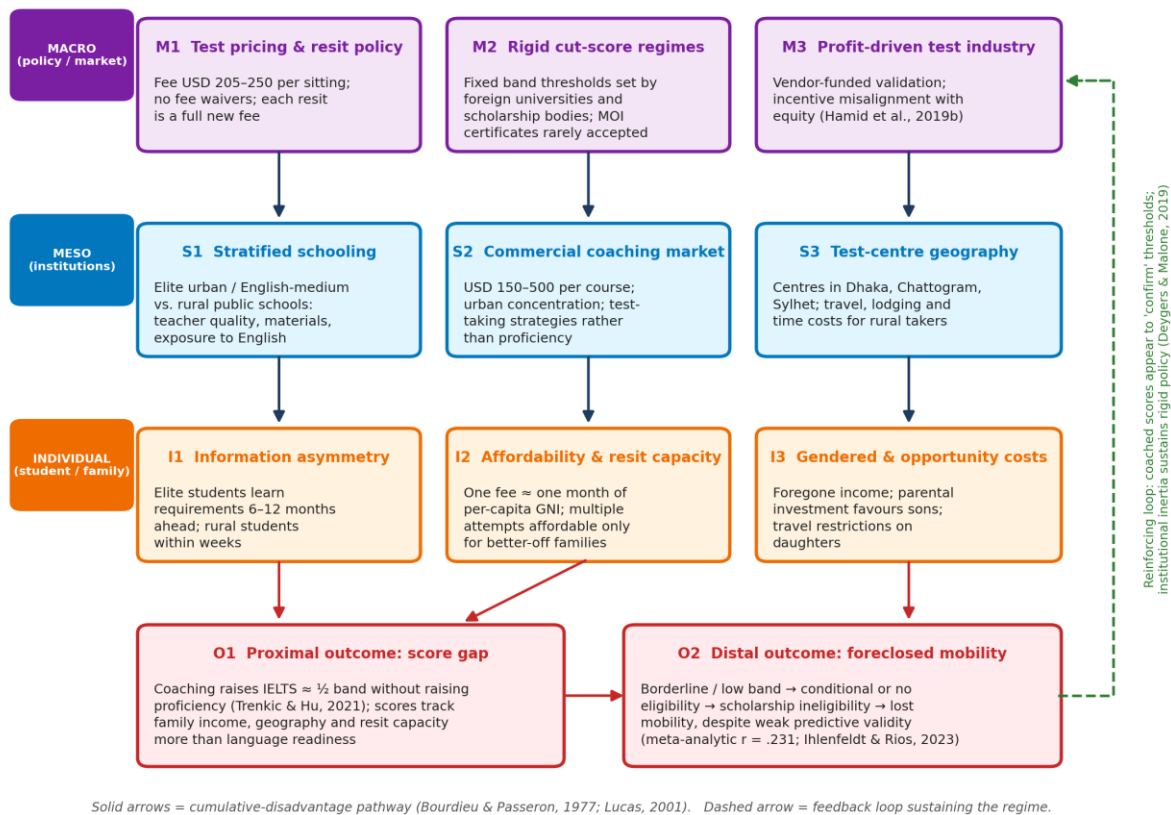
These are not random disadvantages, but a single mechanism, and it's called a stratification lens. When a credential is necessary for mobility, advantaged families get it in whatever way is available, transforming economic capital into cultural capital (Bourdieu & Passeron, 1977); when access to the credential is increased, they continue to maintain their advantage by using qualitative differences in how they obtain the credential, such as coaching that is purchased (Lucas, 2001). Figure 2 illustrates the cumulative-disadvantage pathway from macro conditions to foreclosed mobility, as well as the feedback loop that maintains it, which ELPT in Bangladesh shows both.

Structural inequities were exacerbated by perceptions of test content. Timed, computer-based tests were reported by international graduate students interviewed by Sinclair et al. (2019) to be dehumanizing, and they doubted the relevance of the tasks, including describing bar charts, to the academic English they would need. When considering Kane's (2013) argument about interpretation and use, Deygers et al. (2018) uncovered a significant discrepancy

between the language sampled by the entrance tests and the language needed after admission, thus questioning the fairness and validity-in-use of the tests, not just the item bias.

Figure 2

Cumulative-disadvantage pathway model of inequity in English proficiency testing for Bangladeshi students, synthesized from the included studies



3.3 Financial Barriers and Cost Structures

The most consistent barrier reported was the direct and indirect costs. In Bangladesh, the IELTS and TOEFL registration fee is USD 205 – 250 per sitting, which is approximately one month of per-capita gross national income (World Bank, 2025), and candidates with borderline scores often re-sit to improve their score by half a band, thus increasing the outlay. Coaching programmes, which are mainly in urban centres, are priced at USD 150–500 for 8–12 weeks, and are too expensive for most rural and poor families (Ali & Hamid, 2020; Kh. A. Rahman et al., 2021). Trenkic and Hu (2021) demonstrated that paid coaching resulted in an increase of approximately half a band (e.g., 6.0 to 6.5) but that this increase did not correspond to an increase in the ability to speak English that the IELTS credential is supposed to certify. Hu and Trenkic (2021) also discovered that coached and repeatedly tested candidates were later found to underperform academically in comparison to their IELTS band, which is consistent with Xie's (2013) concern about the validity of scores when candidates are coached and repeatedly tested.

Opportunity costs are greater for low-income students, as preparation time is time away from work; and the opportunity cost of having to re-sit, which several studies refer to, is not just a financial concern, but a concern about effectively maintained inequality, where only those who can afford to re-sit can achieve a score of 7.0 or above on the IELTS.

Table 5

Financial barriers and cost structures identified in the included studies

Barrier	Description	Illustrative evidence	Level(s)
Registration fee	USD 205–250 per sitting (IELTS/TOEFL); DET ≈ USD 70	Hamid & Baldauf (2011); Hamid (2016); Isbell & Kremmel (2020)	Individual; family
Coaching	8–12 weeks; USD 150–500; urban; strategy-focused	Trenkic & Hu (2021): +½ band, no transfer to proficiency	Individual; institutional
Resits	Repeated sittings to reach scholarship cut scores (e.g., 7.0)	Hamid (2016); Hu & Trenkic (2021); weak validity ($r = .231$; Ihlenfeldt & Rios, 2023)	Individual; family
Travel and time	Centres in major cities; lodging, transport, foregone income	Hamid & Baldauf (2011); Terasawa (2025)	Individual; community

3.4 Institutional and Policy-Level Barriers and Enablers

It was identified that there is a lack of transparency in requirements imposed by destination institutions and scholarship programmes, and that they are not well evidence-based. Coney and Isbell (2026) examined ELPT policy at 146 US research-intensive universities and found that, when controlling for institutional characteristics, there was no strong link between higher undergraduate cut scores and timely graduation to support the educational argument for high cut scores. Deygers and Malone (2019) indicated that the admissions officer is not necessarily assessment literate and able to interpret the scores they need, and thresholds are screening conventions and not validated. Hamid et al. (2019b) label this as “linguistic gatekeeping” within a neoliberal profit-making system in which test providers have an interest in maintaining high barriers to success with the test, and have little incentive to make the test easier to pass, as was first described in the case of Australia's language-in-migration policy (Hoang & Hamid, 2017). Both O'Loughlin (2011) and McNamara & Ryan (2011) conclude from a justice perspective that a cut score may be psychometrically fair yet have unfair consequences.

Enablers are in place but not well institutionalized. There is some limited and inconsistent uptake of MOI certificates, DET scores, or direct assessment by some universities. Campbell and Neff's (2020) systematic review of 105 scholarship programmes for students from the Global South revealed that the language requirements disproportionately affect the lower-income candidates and that very few programmes waive or modify the requirements. The evidence from Bangladesh reveals a similar scenario at home: EMI in higher education has benefited those who already know English, and not the others (Sultana, 2014).

3.5 Lived Experiences: Stress, Agency and Gender

Bangladeshi students tell of the charged rite of passage of preparing for ELPT, a failure of which blocks mobility, funding and social status. Test takers in the global samples refer to IELTS as a 'passport' and 'tollgate' and feel that they have to pay for it repeatedly, and there is no alternative (Hamid, 2016; Hamid et al., 2019a; Sinclair et al., 2019). Meanwhile, students are also given the opportunity to be active and use free online resources (such as YouTube and BBC Learning English), peer information networks and ongoing self-study, which are bottom-up approaches that indicate the potential of open-access preparation, but which cannot replace the face-to-face coaching and repeated

efforts that wealthier peers can afford. Gender patterns are not additive, but intersectional: while the average English proficiency is more or less gender neutral (Denies et al., 2022), a number of studies have also found that families invest more in their sons' international aspirations and limit their daughters' access to travel to test centres in cities, so that gender affects access to preparation and sitting rather than performance itself (Abbasian et al., 2020; Hamid & Baldauf, 2011).

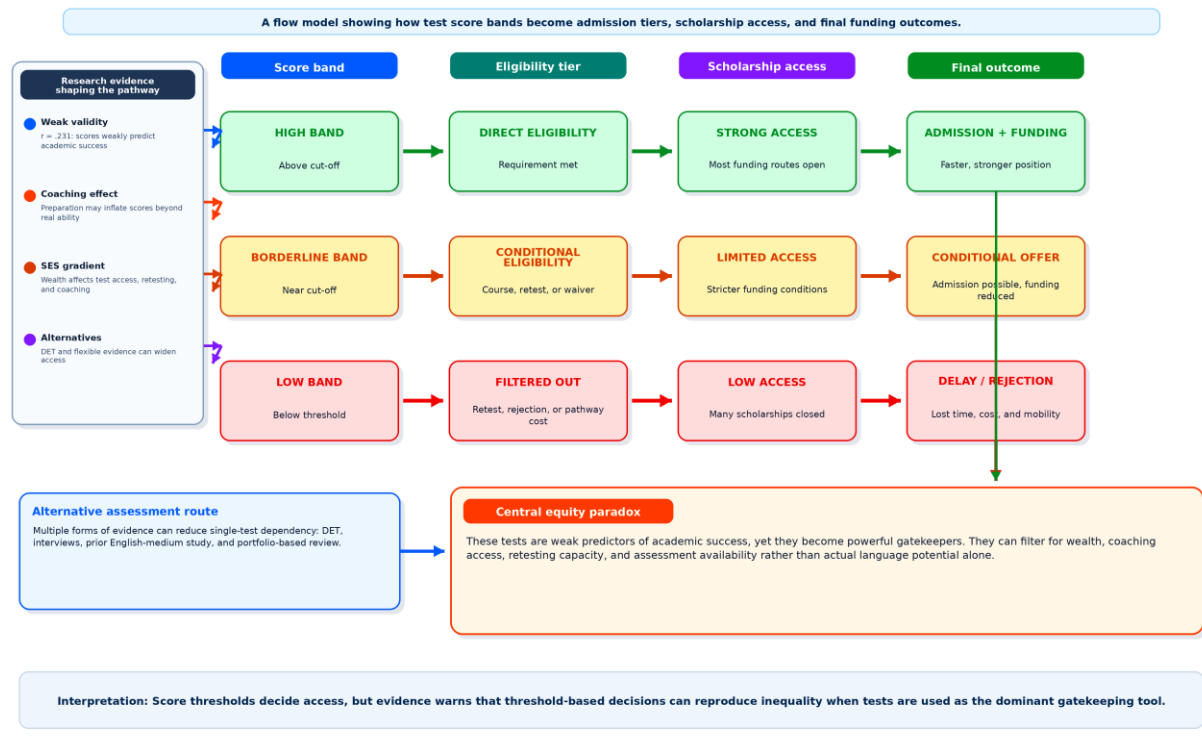
3.6 From Test Scores to Admissions and Scholarships

ELPT scores are not determinative of admissions or scholarship outcomes, but rather mediated (Figure 3). The meta-analysis by Ihlenfeldt and Rios (2023) of 132 effect sizes showed a weak positive relationship between ELPT scores and academic achievement ($r = .231$, $p < .001$), and that there was no significant difference between the IELTS and TOEFL. Institution-level studies report similar magnitudes (Bridgeman et al., 2016; Schoepp, 2018). This loose connection is in stark contrast to the decided gatekeeping function of the tests. Scholarship panels are based on ELPT scores as well as academic record, leadership, and sometimes financial need, but the cut scores are not transparent and are different for different programmes, so applicants are required to make plans under conditions of uncertainty.

There are some alternative tests, which provide partial relief. Isaacs et al. (2023) found that postgraduates who took the DET had a positive but weak relationship with their credit-weighted first-year grade (adjusted $r = .195$) compared to those who were admitted on the basis of traditional tests, and that the DET had a positive but insignificant relationship with the credit-weighted first-year grade of undergraduates (adjusted $r = -.112$), with a lower cost and quicker turnaround than the traditional tests. The negative undergraduate coefficient, although small, is telling, not unusual. The undergraduate sub-sample was smaller ($n = 492$) and was admitted during the pandemic, when weaker students may have been admitted late in the cycle, and DET was the only option; first-year undergraduate grades are less dependent on academic language than postgraduate assessment is, and range restriction above the cut score attenuates any true relationship (Isaacs et al., 2023). This finding thus indicates selection effects and the boundaries of the use of any one score as a predictor, rather than that higher English proficiency has a negative impact on undergraduates. DET is also more accessible than legacy tests, as evidenced by the fact that it extrapolates in a similar manner to university speaking demands (Isbell et al., 2024), despite the concerns about construct-coverage (Wagner, 2020; Yao, 2023).

Figure 3

Schematic pathways from English proficiency test scores to international admissions and scholarship outcomes



4. Discussion

This review was a synthesis of the evidence from 38 studies on equity, access, and barriers in ELPT for Bangladeshi students. The results show that the high-stakes testing of English is a dual process of being a gatekeeper and a product to be marketed: it claims to be an objective assessment of proficiency, and it re-creates the socioeconomic inequalities that give rise to who can demonstrate proficiency in the first place.

4.1 Synthesis of Key Findings

There are three mechanisms of inequity that are at play at different levels. At the individual level, the differential access to fees (USD 205-250 per sitting, which is about a month of per capita income) and the lesser recognition of the cheaper DET (about USD 70) by selective universities is a form of differential access (Isbell & Kremmel, 2020). Preparation is differentiated at the institutional level, with elite urban schools offering intensive, strategy-focused preparation and rural public schools lacking infrastructure and specialised teaching (Ali & Hamid, 2020; Hamid & Baldauf, 2011). Policies that are inflexible in determining scholarship eligibility, and the lack of an alternative route, perpetuate these inequities. Coaching, in addition, is not innocuous: it is correlated with higher IELTS scores, but lower levels of proficiency (Trenkic & Hu, 2021); it is correlated with poorer academic results in the future than band (Hu & Trenkic, 2021), which is particularly relevant in the case of Bangladesh, where commercial coaching has grown and paid preparation is now a prerequisite for admission.

4.2 Theoretical Contribution: Critical Language Testing and the Dual Role of ELPTs

The Bangladeshi evidence is not only illustrative of critical language testing theory, but it also refines the theory. Shohamy (2001) said that the power of tests is in the uses to which they are put, and McNamara and Ryan (2011)

made a distinction between fairness (psychometric properties of the instrument) and justice (defensibility of the social decisions made with the instrument). In the present synthesis, the mechanism of how a fair test can result in unfair decisions in a Global South context is specified. It is the communicative competence that makes ELPT persuasive to institutions because it is indeed an index of communicative competence. However, the factors that can influence the development and expression of competence on test day (schooling, coaching, fees, resit capacity, travel and information) are distributed by class, geography and gender. Thus, a score is a combination of two things: Proficiency and Position. This is the “dual role” of ELPTs: the same instrument that certifies language ability also transforms economic capital into an admissible credential, as the scores of those who buy preparation are increased regardless of their proficiency (Trenkic & Hu, 2021). Institutions that only consider the score as a measure of ability endorse the positional aspect as merit.

Three additional pieces are added below. First, the Bangladeshi case illustrates how the test, which is controlled by a foreign institution, makes migration aspirations dependent on global inequalities, as there is no negotiating with the English requirement (Hamid et al., 2019b); and test takers themselves view the test as a means of earning money rather than a neutral assessment (Hamid et al., 2019a; Pearson, 2019). Secondly, the gatekeeping has created a local market response which is also unequal: coaching centres are of varying quality and provide individual coaching to the privileged and rote memorization for others (Kh. A. Rahman et al., 2021). Third, alternative tests make the front door cheaper to enter, but do not ensure that students are ready to go: DET makes the front door cheaper to enter, but not necessarily ready to go (Isaacs et al., 2023), suggesting that access reforms need to be complemented with academic supports after entry.

4.3 Equity Implications for Bangladeshi Students’ Pathways

Barriers are not incidental but structural: low-income students have to pay a fee, pay for coaching, pay for resits; rural students have to travel to centres and coaching; female students have culturally specific barriers to travelling and investing; public-school graduates have been taught for years without the English they are being tested on (Abbasian et al., 2020). The size of such effects is illustrated by Li et al. (2024): the sequential implementation of an English listening test in China's college entrance exam reduced the percentile rank of rural students compared to urban students, widened the rural–urban college gap by 30%, and forced over 54,000 rural students to give up college spots. The causal mechanism, a requirement for English which exacerbates pre-existing geographic and socioeconomic inequality, is easily translated to Bangladesh, although institutional details vary. For the policy on scholarships, the implication is clear: Fixed cut scores are not a neutral filter where band scores are more closely related to family income than to learning readiness. Table 6 provides an overview of the mechanisms, enablers and leverage points found at each level.

Table 6

Equity-relevant mechanisms, enablers and leverage points identified across levels

Level	Mechanisms of inequity	Mechanisms of facilitation	Evidence	Leverage points
Individual	Fees; coaching cost; resits; time constraints; gendered investment	Fee waivers; DET; open-access preparation; flexible scheduling; peer mentoring	Isbell & Kremmel (2020); Abbasian et al. (2020); Yao (2023)	Targeted subsidies for rural and low-income students; transparent communication of affordable options
Institutional	Elite–public coaching gap; teacher qualification	Government-funded test-preparation; teacher development;	Ali & Hamid (2020); Ali et al. (2020); Kh.	Test literacy in the national curriculum; capacity building for

	variance; EMI disparities; rural centre scarcity	remote coaching; quality standards for coaching centres	A. Rahman et al. (2021); Sultana (2014)	rural schools; university–school assessment pilots
Policy	Rigid cut scores; non- recognition of MOI evidence; foreign- controlled admissions policy; vendor-funded validation	Validated alternatives; flexible band policies; transparent scholarship criteria; bilateral recognition agreements	Deygers & Malone (2019); Hamid et al. (2019b); Coney & Isbell (2026); Li et al. (2024)	Ministry–university dialogue on recognition; concordance studies; independent validation with South Asian samples

4.4 Implications for Policy, Practice and Test Design

The evidence results in four sets of actionable recommendations. For the Bangladesh Ministry of Education and the University Grants Commission: (1) create a test-fee support fund for rural, low-income and first-generation test takers, similar to the test fee support fund in India and Singapore; (2) integrate test literacy (what the test measures, what alternatives are available, how to prepare without commercial coaching) into the Higher secondary curriculum; and (3) negotiate bilateral recognition of Bangladeshi EMI qualifications with key destination systems. For international universities and scholarship providers: (4) use flexible band thresholds, rather than single cut scores, for evidence that a borderline score (e.g., 0.5 band below) can be used as a basis for conditional admission with in-session support, as evidence suggests that cut scores do not predict graduation (Coney & Isbell, 2026); (5) formally recognize MOI certificates from accredited English-medium institutions and other validated alternatives (e.g., DET); and (6) publish eligibility criteria so that applicants can plan. For test providers and coaching centres: (7) implement transparent fee and pricing policies, provide remote delivery for underserved areas, fund independent, non-vendor validation of test samples using South Asian samples; and (8) move away from test-specific drilling and towards evidence-based pedagogy that builds proficiency, not scores. These measures aim to address the core injustice identified in this report: that today's high-stakes tests have the effect of reinforcing and widening social inequalities, and are not simply tests of language proficiency.

5. Conclusions

5.1 Main Conclusions and the “So What?”

This review, based on 38 studies, reveals that ELPTs are used as a measure of communicative competence as well as a strong economic and social filter for Bangladeshi students. The differences in outcomes are not as much due to the design of the tests themselves, but to the interplay between the costs of the tests, information asymmetries, and the geographically concentrated coaching and rigid institutional thresholds (Hu & Trenkic, 2021; Xie, 2013). While the tests do not have a psychometric bias against low-income, rural, or female students, they are unfair in their application because the opportunities to pass the test are not equally available and the score differences are interpreted as signs of merit. The policy message is that the English requirements are an 'inequality multiplier' which can only be offset by subsidy and recognition reforms. For universities, it is screening conventions, not standards, that are the cut scores without concordance and support. That a set IELTS cut-off might systematically exclude those it is intended to reach is an issue for scholarship bodies that are about widening opportunity.

5.2 A Forward-Looking Agenda

There are two avenues of reform that can be achieved. The first is recognition reform, which is supported by the included validity studies, such as flexible band thresholds, accepting MOI and portfolio evidence, and conditional admission based on concordance. The second is digital and open access: remote testing, DET-style low-cost formats,

and open-access preparation can help to break down fee, travel, and coaching barriers, as long as it is independently validated with Global South samples, and combined with post-entry support. Research should now (1) take a longitudinal approach to the scholarship and degree outcomes of students admitted through alternative assessments; (2) test the predictive validity of DET with other cohorts of students from Bangladesh and South Asia, outside of the vendor-funding context; (3) examine the effects of coaching beyond test scores, and explore academic trajectories; (4) undertake institutional ethnographies of coaching centres; and (5) pursue intersectional designs, which are under-represented in the literature.

5.3 Limitations

The results need to be interpreted in the context of four limitations. First, the main literature on Bangladesh is urban and institutional – most of the studies are about Dhaka, Chattogram and elite institutions, with experiences of rural and government schools under-represented and barriers possibly underestimated. Second, only eight studies were Bangladesh-specific; the rest of the evidence was from comparator contexts based on transferable mechanisms, and transfer is a judgment. Third, the voices of students are heard more than those of institutional actors, meaning that structural barriers are reported as perceived, and few studies report results by gender, class, or type of school. Fourth, evidence of predictive validity is based on samples other than Bangladeshi; and the use of Google Scholar (which has a capped number of results and does not provide transparency of indexing) and the lack of PROSPERO registration are methodological limitations noted above (Haddaway et al., 2015). Despite these limitations, the review provides a multi-level, theoretically framed synthesis that places Bangladesh in the context of gatekeeping scholarship in the world and provides an agenda for reform for policy makers, universities and scholarship providers.

Declarations

Acknowledgements

None.

Competing Interests

None.

Ethical Approval

This study was granted an exemption from requiring ethics approval as it does not involve human participants or the collection of sensitive personal data. The research is based on bibliometric methods, utilising secondary data exclusively from different databases. As such, it adheres to institutional guidelines that classify this type of study as low-risk and not subject to formal ethics approval.

Author Contributions

Auhtor¹: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft

Auhtor²: Methodology, Supervision, Investigation, review and editing

References

- Abbasian, R., Hadian, B., & Vaez-Dalili, M. (2020). Examination of the role of family socio-economic status and parental education in predicting English as a foreign language learners' receptive skills performance. *Cogent Education*, 7(1), 1710989. <https://doi.org/10.1080/2331186X.2019.1710989>
- Ali, Md. M., & Hamid, M. O. (2020). Teaching English to the test: Why does negative washback exist within secondary education in Bangladesh? *Language Assessment Quarterly*, 17(2), 129–146. <https://doi.org/10.1080/15434303.2020.1717495>

- Ali, Md. M., Hamid, M. O., & Hardy, I. (2020). Ritualisation of testing: Problematising high-stakes English-language testing in Bangladesh. *Compare: A Journal of Comparative and International Education*, 50(4), 533–553. <https://doi.org/10.1080/03057925.2018.1535890>
- Bourdieu, P., & Passeron, J.-C. (1977). *Reproduction in education, society and culture* (R. Nice, Trans.). Sage.
- Bridgeman, B., Cho, Y., & DiPietro, S. (2016). Predicting grades from an English language assessment: The importance of peeling the onion. *Language Testing*, 33(3), 307–318. <https://doi.org/10.1177/0265532215583066>
- Campbell, A. C., & Neff, E. (2020). A systematic review of international higher education scholarships for students from the Global South. *Review of Educational Research*, 90(6), 824–861. <https://doi.org/10.3102/0034654320947783>
- Cardwell, R. L., Nydick, S. W., Lockwood, J. R., & von Davier, A. A. (2024). Practical considerations when building concordances between English tests. *Language Testing*, 41(1), 192–202. <https://doi.org/10.1177/02655322231195027>
- Coney, N., & Isbell, D. R. (2026). A survey of English language proficiency tests in international student admissions at US research-intensive universities. *Language Testing*, 43(2), 233–252. <https://doi.org/10.1177/02655322251348617>
- Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research*, 22(10), 1435–1443. <https://doi.org/10.1177/1049732312452938>
- Critical Appraisal Skills Programme. (2024). *CASP checklists*. <https://casp-uk.net/casp-tools-checklists/>
- Denies, K., Heyvaert, L., Dockx, J., & Janssen, R. (2022). Mapping and explaining the gender gap in students' second language proficiency across skills, countries and languages. *Learning and Instruction*, 80, 101618. <https://doi.org/10.1016/j.learninstruc.2022.101618>
- Deygers, B., & Malone, M. E. (2019). Language assessment literacy in university admission policies, or the dialogue that isn't. *Language Testing*, 36(3), 347–368. <https://doi.org/10.1177/0265532219826390>
- Deygers, B., Van den Branden, K., & Van Gorp, K. (2018). University entrance language tests: A matter of justice. *Language Testing*, 35(4), 449–476. <https://doi.org/10.1177/0265532217706196>
- Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLoS ONE*, 10(9), e0138237. <https://doi.org/10.1371/journal.pone.0138237>
- Hamid, M. O. (2016). Policies of global English tests: Test-takers' perspectives on the IELTS retake policy. *Discourse: Studies in the Cultural Politics of Education*, 37(3), 472–487. <https://doi.org/10.1080/01596306.2015.1061978>
- Hamid, M. O., & Baldauf, R. B. (2011). English and socio-economic disadvantage: Learner voices from rural Bangladesh. *Language Learning Journal*, 39(2), 201–217. <https://doi.org/10.1080/09571736.2011.573687>
- Hamid, M. O., & Erling, E. J. (2016). English-in-education policy and planning in Bangladesh: A critical examination. In R. Kirkpatrick (Ed.), *English language education policy in Asia* (pp. 25–48). Springer. https://doi.org/10.1007/978-3-319-22464-0_2
- Hamid, M. O., Hardy, I., & Reyes, V. (2019a). Test-takers' perspectives on a global test of English: Questions of fairness, justice and validity. *Language Testing in Asia*, 9, 16. <https://doi.org/10.1186/s40468-019-0092-9>
- Hamid, M. O., Hoang, N. T. H., & Kirkpatrick, A. (2019b). Language tests, linguistic gatekeeping and global mobility. *Current Issues in Language Planning*, 20(3), 226–244. <https://doi.org/10.1080/14664208.2018.1495371>
- Hoang, N. T. H., & Hamid, M. O. (2017). "A fair go for all?" Australia's language-in-migration policy. *Discourse: Studies in the Cultural Politics of Education*, 38(6), 836–850. <https://doi.org/10.1080/01596306.2016.1199527>

- Hu, R., & Trenkic, D. (2021). The effects of coaching and repeated test-taking on Chinese candidates' IELTS scores, their English proficiency, and subsequent academic achievement. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1486–1501. <https://doi.org/10.1080/13670050.2019.1691498>
- Ihlenfeldt, S. D., & Rios, J. A. (2023). A meta-analysis on the predictive validity of English language proficiency assessments for college admissions. *Language Testing*, 40(2), 276–299. <https://doi.org/10.1177/02655322221112364>
- Isaacs, T., Hu, R., Trenkic, D., & Varga, J. (2023). Examining the predictive validity of the Duolingo English Test: Evidence from a major UK university. *Language Testing*, 40(3), 748–770. <https://doi.org/10.1177/02655322231158550>
- Isbell, D. R., Crowther, D., & Nishizawa, H. (2024). Speaking performances, stakeholder perceptions, and test scores: Extrapolating from the Duolingo English Test to the university. *Language Testing*, 41(2), 233–262. <https://doi.org/10.1177/02655322231165984>
- Isbell, D. R., & Kremmel, B. (2020). Test review: Current options in at-home language proficiency tests for making high-stakes decisions. *Language Testing*, 37(4), 600–619. <https://doi.org/10.1177/0265532220943483>
- Islam, M. S., & Stapa, M. B. (2021). Students' low proficiency in spoken English in private universities in Bangladesh: Reasons and remedies. *Language Testing in Asia*, 11, 22. <https://doi.org/10.1186/s40468-021-00139-0>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Li, H., Meng, L., Mu, K., & Wang, S. (2024). English language requirement and educational inequality: Evidence from 16 million college applicants in China. *Journal of Development Economics*, 168, 103271. <https://doi.org/10.1016/j.jdeveco.2024.103271>
- Long, H. A., French, D. P., & Brooks, J. M. (2020). Optimising the value of the Critical Appraisal Skills Programme (CASP) tool for quality appraisal in qualitative evidence synthesis. *Research Methods in Medicine & Health Sciences*, 1(1), 31–42. <https://doi.org/10.1177/2632084320947559>
- Lucas, S. R. (2001). Effectively maintained inequality: Education transitions, track mobility, and social background effects. *American Journal of Sociology*, 106(6), 1642–1690. <https://doi.org/10.1086/321300>
- Maeda, Y., Caskurlu, S., Kenney, R. H., Kozan, K., & Richardson, J. C. (2022). Moving qualitative synthesis research forward in education: A methodological systematic review. *Educational Research Review*, 35, 100424. <https://doi.org/10.1016/j.edurev.2021.100424>
- McNamara, T., & Ryan, K. (2011). Fairness versus justice in language testing: The place of English literacy in the Australian Citizenship Test. *Language Assessment Quarterly*, 8(2), 161–178. <https://doi.org/10.1080/15434303.2011.565438>
- O'Loughlin, K. (2011). The interpretation and use of proficiency test scores in university selection: How valid and ethical are they? *Language Assessment Quarterly*, 8(2), 146–160. <https://doi.org/10.1080/15434303.2011.564698>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Pearson, W. S. (2019). Critical perspectives on the IELTS test. *ELT Journal*, 73(2), 197–206. <https://doi.org/10.1093/elt/ccz006>
- Pill, J., Frost, K., & Macqueen, S. (2025). Fairness and justice in language testing: The challenge of Tim McNamara's legacy. *Language Testing*, 42(1), 124–134. <https://doi.org/10.1177/02655322241277268>
- Rahman, Kh. A., Seraj, P. M. I., Hasan, Md. K., Namaziandost, E., & Tilwani, S. A. (2021). Washback of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia*, 11, 12. <https://doi.org/10.1186/s40468-021-00129-2>

- Rahman, M. M., Islam, M. S., Karim, A., Chowdhury, T. A., Rahman, M. M., Seraj, P. M. I., & Singh, M. K. M. (2019). English language teaching in Bangladesh today: Issues, outcomes and implications. *Language Testing in Asia*, 9, 9. <https://doi.org/10.1186/s40468-019-0085-8>
- Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in Nursing & Health*, 23(4), 334–340. [https://doi.org/10.1002/1098-240X\(200008\)23:4<334::AID-NUR9>3.0.CO;2-G](https://doi.org/10.1002/1098-240X(200008)23:4<334::AID-NUR9>3.0.CO;2-G)
- Schoepp, K. (2018). Predictive validity of the IELTS in an English as a medium of instruction environment. *Higher Education Quarterly*, 72(4), 271–285. <https://doi.org/10.1111/hequ.12163>
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Longman.
- Shohamy, E. (2022). Critical language testing, multilingualism and social justice. *TESOL Quarterly*, 56(4), 1445–1457. <https://doi.org/10.1002/tesq.3185>
- Sinclair, J., Larson, E. J., & Rajendram, S. (2019). “Be a machine”: International graduate students’ narratives around high-stakes English tests. *Language Assessment Quarterly*, 16(2), 236–252. <https://doi.org/10.1080/15434303.2019.1628238>
- Sultana, S. (2014). English as a medium of instruction in Bangladesh’s higher education: Empowering or disadvantaging students? *Asian EFL Journal*, 16(1), 11–52.
- Terasawa, T. (2025). Relationship between English proficiency and socioeconomic status in Asia: Quantitative cross-national analysis. *World Englishes*, 44(3), 492–516. <https://doi.org/10.1111/weng.12705>
- Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, 45. <https://doi.org/10.1186/1471-2288-8-45>
- Tong, A., Flemming, K., McInnes, E., Oliver, S., & Craig, J. (2012). Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology*, 12, 181. <https://doi.org/10.1186/1471-2288-12-181>
- Trenkic, D., & Hu, R. (2021). Teaching to the test: The effects of coaching on English-proficiency scores for university entry. *Journal of the European Second Language Association*, 5(1), 1–15. <https://doi.org/10.22599/jesla.74>
- Wagner, E. (2020). Duolingo English Test, revised version July 2019. *Language Assessment Quarterly*, 17(3), 300–315. <https://doi.org/10.1080/15434303.2020.1771343>
- World Bank. (2025). *GNI per capita, Atlas method (current US\$) – Bangladesh* [Data set]. World Development Indicators. <https://data.worldbank.org/indicator/NY.GNP.PCAP.CD?locations=BD>
- Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly*, 10(2), 196–218. <https://doi.org/10.1080/15434303.2012.721423>
- Yao, D. (2023). Examining the subjective fairness of at-home and online tests: Taking Duolingo English Test as an example. *PLoS ONE*, 18(9), e0291629. <https://doi.org/10.1371/journal.pone.0291629>

Supplementary Material

Table S1. Characteristics and quality rating of the 38 included studies (listed alphabetically; CASP bands: High ≥ 14, Moderate 10–13, Low < 10)

#	Study	Context / sample	Design	Test(s) examined	Equity-relevant focus	CASP band
1	Abbasian et al. (2020)	Iran	Quantitative	General EFL (receptive skills)	SES, parental education and proficiency	High
2	Ali & Hamid (2020)	Bangladesh	Qualitative	HSC English; IELTS prep	Negative washback; teaching to the test	High
3	Ali, Hamid & Hardy (2020)	Bangladesh	Qualitative	HSC English; IELTS	Ritualisation of high-stakes testing	High
4	Bridgeman, Cho & DiPietro (2016)	USA	Quantitative	TOEFL iBT	Predictive validity of grades	High
5	Campbell & Neff (2020)	Global South (105 programmes)	Systematic review	IELTS; TOEFL (requirements)	Scholarship language requirements as barrier	High
6	Cardwell et al. (2024)	International	Quantitative	IELTS; TOEFL; DET; PTE	Concordance between tests	High
7	Coney & Isbell (2026)	USA (146 universities)	Quantitative (policy survey)	IELTS; TOEFL; DET; PTE	Cut-score policy and graduation	High
8	Denies et al. (2022)	Europe (cross-national)	Quantitative	Standardised L2 tests	Gender gap in L2 proficiency	High
9	Deygers, Van den Branden & Van Gorp (2018)	Belgium	Mixed methods	University entrance tests	Justice and validity-in-use	High
10	Deygers & Malone (2019)	Belgium; USA	Qualitative	IELTS; TOEFL (admissions use)	Assessment literacy of admissions staff	High
11	Hamid (2016)	Australia (international test-takers)	Qualitative	IELTS	Retake policy; cost burden	Moderate
12	Hamid & Baldauf (2011)	Bangladesh (rural)	Qualitative	General English; IELTS/TOEFL aspirations	Socioeconomic disadvantage	High
13	Hamid & Erling (2016)	Bangladesh	Critical policy analysis	General English	English-in-education policy and capacity	Moderate
14	Hamid, Hardy & Reyes (2019a)	Global (49 countries)	Mixed methods (survey)	IELTS	Fairness, justice and validity perceptions	Moderate

#	Study	Context / sample	Design	Test(s) examined	Equity-relevant focus	CASP band
15	Hamid, Hoang & Kirkpatrick (2019b)	Global	Critical policy analysis	IELTS; TOEFL; PTE	Linguistic gatekeeping and mobility	High
16	Hoang & Hamid (2017)	Australia	Critical policy analysis	IELTS	Language-in-migration policy	Moderate
17	Hu & Trenkic (2021)	China; UK	Quantitative	IELTS	Coaching, repeat testing and achievement	High
18	Ihlenfeldt & Rios (2023)	International (k = 132)	Meta-analysis	IELTS; TOEFL; PTE	Predictive validity	High
19	Isaacs, Hu, Trenkic & Varga (2023)	UK (n = 1,881)	Quantitative	DET; IELTS; TOEFL	Predictive validity of DET	High
20	Isbell, Crowther & Nishizawa (2024)	USA	Mixed methods	DET	Extrapolation to university speaking	High
21	Isbell & Kremmel (2020)	International	Test review	DET; TOEFL Home; IELTS Indicator	At-home test options and access	Moderate
22	Islam & Stapa (2021)	Bangladesh (private universities)	Qualitative	Spoken English (general)	Low spoken proficiency; remedies	Low
23	Li, Meng, Mu & Wang (2024)	China (16 million applicants)	Quantitative (natural experiment)	NCEE English listening	English requirement and rural–urban gap	High
24	McNamara & Ryan (2011)	Australia	Critical policy analysis	Citizenship English test	Fairness versus justice	High
25	O’Loughlin (2011)	Australia	Critical policy analysis	IELTS; TOEFL	Ethics of cut-score use	Moderate
26	Pearson (2019)	International	Critical analysis	IELTS	Critical perspectives on IELTS	Moderate
27	Pill, Frost & Macqueen (2025)	International	Critical conceptual analysis	ELPTs (general)	Fairness and justice frameworks	High
28	Rahman et al. (2019)	Bangladesh	Review	General ELT	ELT issues and outcomes	Moderate
29	Rahman, Seraj et al. (2021)	Bangladesh (secondary)	Qualitative	SSC/HSC English; IELTS prep	Washback on teaching and learning	Moderate
30	Schoepp (2018)	UAE	Quantitative	IELTS	Predictive validity in EMI	High
31	Shohamy (2022)	International	Critical conceptual analysis	ELPTs (general)	Critical language testing and justice	Moderate

#	Study	Context / sample	Design	Test(s) examined	Equity-relevant focus	CASP band
32	Sinclair, Larson & Rajendram (2019)	Canada (international graduate students)	Qualitative	IELTS; TOEFL	Lived experience; dehumanisation	High
33	Sultana (2014)	Bangladesh (higher education)	Qualitative	EMI (general English)	EMI empowering or disadvantaging	Low
34	Terasawa (2025)	Asia (30 jurisdictions)	Quantitative	Proficiency indices	SES–English gradient	High
35	Trenkic & Hu (2021)	China	Quantitative (experimental)	IELTS; DET; independent measures	Coaching effects on scores	High
36	Wagner (2020)	International	Test review	DET	Construct coverage of DET	Moderate
37	Xie (2013)	China	Quantitative	CET; IELTS prep	Test preparation and score validity	High
38	Yao (2023)	China	Quantitative (survey)	DET	Subjective fairness of at-home testing	Moderate

DET = Duolingo English Test; EMI = English-medium instruction; HSC = Higher Secondary Certificate; NCEE = National College Entrance Examination (China); SES = socioeconomic status.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of RISE and/or the editor(s). RISE and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.