"We would like to see ourselves in the test:" The views of Francophone West African English learners as potential test takers of the Duolingo English Test (DET)

Kadidja Koné

Ecole Normale d'Enseignement Technique et Professionnel, Mali

Paula Winke

Michigan State University, USA

Matt Gordon

Michigan State University, USA

High-stakes test design must be informed by test takers' views so that the test takers can benefit from the tests (Fox & Cheng, 2015; Hamid et al., 2019; Jin, 2023; Kang, Miao, & Hirschi, 2024; O'Sullivan, 2012). High-stakes English tests are being administered in Francophone West Africa, yet little research has investigated test takers' experiences with them. Thus, we investigated 64 Malian test takers' views after taking a standardized English test, the Duolingo English Test (DET). Across 10 afternoon focus groups, we asked test takers about their perceptions of the test (which they had taken in the morning) and what they believed designers should do to make the test better for them. Emergent themes, found through analyses using MAXQDA, centered on the test's features/areas, technology, inclusivity, and professionalization opportunities. Test takers additionally discussed the test-taking anxieties they experienced and suggested recommendations for the test designers. In sum, local culture, infrastructure, and educational contexts interplayed with exam experiences. Test takers expressed a desire for more African-centric exam content. Overall, the test

Email address for correspondence: kadidja_kone@yahoo.fr

© The Author(s) 2025. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits the user to copy, distribute, and transmit the work provided that the original authors and source are credited. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

takers enjoyed participating, yet discussed that for them, the test would be more accessible if it cost less, was mobile-phone based, and could be paid for through local banking systems.

Keywords: Duolingo English Test, high-stakes tests, test-taking experiences, Francophone West Africa, technology

Introduction

In the fall of 2012, the first author of this study, Kadidja Koné, was applying for a United States Fulbright scholarship to obtain a master's degree in applied linguistics in the United States. As part of the requirement for applying for the Fulbright, she had to take the Test of English as a Foreign Language internet-based test (TOEFL iBT) at a test center in Bamako, Mali. Because she had not yet been to the United States, some test content was unfamiliar, which made the exam difficult for her, even though she believed she was proficient in English, having learned English for over 16 years, and, at that time, she was a teacher of English in the upper-level of a public secondary school. Later, when focusing on language assessment research as an assistant professor, English language testing was a discussion point Kadidja Koné had with other language assessment researchers, including the second author on this paper, Paula Winke. Are standardized, international English language tests appropriate for the Francophone West African context? More viewpoints on this issue were clearly needed, and the most important viewpoints to obtain would be from more Francophone West African English-language learners who had the opportunity to take a standardized English language test.

Such research would be timely. The *New York Times* (Walsh, 2023) has predicted that population growth and economic changes in Africa will transform African countries, and also affect Africa's relationship with the rest of the world. To quote the *Times*, "The world is becoming more African" (p. 4), stressing Africa's influence on world culture, economics, health, and politics. As McDermott (2025, p. 14) observed, "Africa's demographic boom has led to the idea that the 21st is the African century." Meanwhile, Africa is learning English for international purposes. For example, in Francophone West Africa (e.g., Benin, Burkina Faso, Mali, Niger, and Senegal), pupils start learning English in the public

schools at the age of 12. At the same time, the African educational context, specifically Francophone West Africa, is under-studied and underrepresented within research on high-stakes, English-language proficiency tests (ELPTs). One possible set of exceptions here is the studies undertaken in South Africa in respect of the accessibility and accountability for high-stakes tests of academic literacy (Rambiritch, 2014, 2015; Van der Walt & Steyn, 2007). In this study, however, we empirically investigate specifically Francophone West African test takers' experiences and reflections on taking a standardized, international English test in Africa.

Literature review

Not knowing how test takers in Francophone West Africa view high-stakes ELPTs and how the tests affect them as learners is a blind spot for designers and administrators of ELPTs in that location, especially because "language testers are shifting towards uses and consequences of tests in terms of their impact for individuals, education and societies" (Shohamy, 2022, p. 1454).

The theoretical importance of collecting test takers' views

Language testing theorists have suggested that test takers' views are an important part of *validation research* (Cheng & DeLuca, 2011; Jin, 2023), that is, collecting evidence to determine whether test-score interpretations can be trusted (Kane, 2013). Test takers' views are additionally described as potentially important for test design and test-revision purposes. For example, O'Sullivan (2012) has suggested that if a high-stakes test is designed with the test takers' needs, characteristics, and contexts in mind, the test will represent the test takers' learning and targeted-language use *and* enhance the test's score-interpretation validity. In other words, gathering test takers' views can help boost *quality* in large-scale language assessments and is part of a *participatory approach* to test development, according to Little and Erickson (2015), who wrote:

> [S]tudents are undoubtedly the largest group affected by testing and assessment. The value of actively involving students has been pointed out by many researchers in the language assessment field … and the decision to do so can easily be justified in terms of

> validity in a wide sense that embraces democratic principles, ethics, and potential pedagogical impact (p. 134).

Jin (2023) recently expanded on the reasons why test-taker input should be taken into consideration. She noted that by interactively engaging test takers (that is, by coming to understand their views on tests), test developers can do more than monitor the validity of test score interpretation and improve the quality of their tests' designs. Additionally, they can harness more public trust in their tests' score uses and can empower the test takers themselves (Rambiritch, 2014, 2015) by letting them know that they are active agents participating in the test design process, which can help with test taker motivation and even improve test taker performance. Jin stressed that test takers, when asked their views, provide unique *critical* perspectives on test-score validation and interpretation research. According to Shohamy (2022), critical perspectives provide "attention to the impact of tests on people" (p. 1448) and can also expose social consequences and local test-score uses which may be different from the test designers' intended test-score uses. Additionally, as suggested by Hamid and Huong (2018), collecting and using test taker perspectives on high-stakes language tests is an imperative to humanize them (2018, p. 2):

> If the growing interest in test takers and their perspectives is motivated by a desire to respect and act on their voices, we may need to consider them not just as test-taking subjects but as human beings taking language tests in the pursuit of human desires in a globalised world.

Test takers' views on English-language proficiency tests

To help inform our current study, we reviewed ten empirical studies (Cheng & DeLuca, 2011; Elder et al., 2002; Fox & Cheng, 2015; Hamid & Hoang, 2018; Hamid et al., 2019; Kang, Miao, & Hirschi, 2024; Kim, 2017; Noubandegani, 2012; Sinclair et al., 2019; Winke et al., 2018) that we found (a) were published in peer-reviewed journals and (b) whose authors researched test takers' views regarding an English-language proficiency test. We summarized the ten studies in Table 1 by publication date, so that readers could have a bird's eye view of the trajectory of the research that has taken place over the last 25 years. We noticed that over time, researchers held approximately four phases of thought on the value of collecting test takers' perceptions: First, researchers concluded

that test takers' perceptions on English language proficiency tests were not to be relied on too heavily because their responses were seemingly erratic (Elder et al., 2002). Second, researchers moved to suggesting that test takers' views could be helpful for test design (Cheng & DeLuca, 2011; Noubandegani, 2012). Third, researchers pointed out that test takers are uniquely positioned to add insights on test validation, and thus their perceptions are highly important for test design and for understanding operational testing (Fox & Cheng, 2015; Kim, 2017; Winke et al., 2018). And finally, researchers suggested that collecting test takers' views is essential and that test takers' perceptions must be taken into consideration for the test scores to be fairly and justly interpreted (Hamid & Hoang, 2018; Kunnan, 2008; Sinclair et al., 2019).

Table 1. Ten empirical studies on test takers' perceptions of a recently-taken English proficiency test, in chronological order

Author(s), Year, Journal	English test under investigation	Participants	Perception-elicitation method	Time between test & viewpoint gathering	Analyses	Result details	Main conclusion
1. Elder et al., 2002, *Language Testing*	Simulations of the Test of Spoken English (TSE) speaking tasks	201 international college students from at least three Asian countries, in ESL classes in Australia to prepare for university entrance or already enrolled, took a simulation of a standardized speaking test, with the tasks manipulated for difficulty	After each task, the students took a questionnaire on their perceptions of task difficulty, their familiarity with the task type, & their attitudes towards it.	Questionnaires were administered immediately after each task	Item Response Theory (IRT), t tests on the questionnaire Likert-scale data	Test takers perceptions of task difficulty did not align with hypothesized levels of task difficulty. Test takers found tasks with pictures more enjoyable than those without.	Test takers have insights that can inform test development, but also, participants' erratic perception patterns suggested test designers should not rely too heavily on test taker feedback.
2. Cheng & DeLuca, 2011, *Educational Assessment*	Various	59 undergrad & grad students in English language assessment courses at an Asian English-medium university	Asked to write 300 to 500-word essay about their English-language test-taking experiences in the past	Not controlled; undetermined	Deductive, thematic coding, 2 "raters" [unidentified], IRR = 95%	Test takers provided views on construct representation, construct-irrelevant variance, & test consequences, all part of test validation research.	By acknowledging test takers' experiences, test designers can improve test procedures to ensure score interpretations are accurate.

3. Noubandegani, 2012, *Language Testing in Asia*	TOEFL iBT	100 Iranian university-level English learners who took the TOEFL iBT at their university	Learners took a Likert-scale questionnaire after taking the TOEFL iBT	Undetermined; time between test event & questionnaire not given	Descriptive statistics, factor analyses, & correlations	Students had difficulties typing & using the computer mouse, but enjoyed being allowed to test with more flexible testing time slots.	Learners' perceptions leading into a test may affect their performance. Researching perceptions can help designers remove weak points and improve the test.
4. Fox & Cheng, 2015, *TESL Canada Journal*	TOEFL iBT & others	Two international students in Canada who had taken the TOEFL iBT before admittance; 375 English for Academic Purposes (EAP) students at the same university preparing for admission-testing	The two were interviewed; the 375 took a survey	Not controlled; were asked about past and/or future test taking	Constructivist grounded theory approach, 2nd set of raters re-rated subset: Cronbach's alpha .86; descriptive stats on closed questionnaire items	Test takers voiced concerns about test speededness, test anxiety, & test preparation. They discussed the computer-based reading & writing tasks as more difficult than familiar paper-based tests.	Drawing on test taker insights increases understanding of operational testing. Their views may at times contradict research conclusions or expert judgements.
5. Kim, 2017, *Assessing Writing*	TOEFL writing component	The data comprised 476 postings, but the number of participants and further information about them, except that they were Korean English learners, is unknown.	Researched 150 pages of content written in two online TOEFL-focused forums by Korean English learners who took or were preparing for the TOEFL.	Not controlled; undetermined	A reiterative process of coding, categorizing, & recording within software NVivo. "To improve rigor, the entire data were re-coded after three months by the researcher" (p. 3).	Test takers shared memorization strategies they believed were important & discussed confusion over contradicting outcomes on multiple writing tests & attributed disparities to scoring inconsistencies.	Observing perspectives revealed discrepancies between test taker and test-designer beliefs about what the writing test measures, which called into question the usefulness of test scores.

6. Winke et al., 2018, *TESOL Quarterly*	Cambridge YLTE	19 children aged 7 to 9 took sample versions of the Young Learners Tests of English, Bronze & Silver	Each child was interviewed shortly after taking each test	Interviews were given after a short break (~10 minutes) after testing	Interview transcripts coded in NVivo using inductive approach with emergent themes; classical test theory applied to test scores	Incorrect responses were sometimes due to test method, which was more cognitively complex than their age-related cognitive development level could handle.	Young test takers' views can shed light on a test's relevant & irrelevant constructs that are being measured by the test.
7. Hamid & Hoang, 2018, *TESL-EJ*	IELTS & TOEFL	They reported on the study from Hamid et al. (2019) & described another (second) study: 518 prior test takers: 377 had taken IELTS; 141 had taken TOEFL	The 518 took the survey in Hamid et al. (2019); 260 gave open comments, 28 were in semi-structured interviews, 10 were in three focus groups.	Not controlled; Surveys & other instruments were administered at some point after the actual test event	Qualitative data "were examined for instances of humanising" across seven areas (p. 6).	Test takers' discussed how they would like a more friendly & real-world test that is responsive to their needs. They would like to have less stress when preparing for or taking the TOEFL or IELTS.	Test takers' views should not just be heard, but also acted upon by test designers in a global world.
8. Sinclair et al., 2019, *Language Assessment Quarterly*	IELTS	Four international graduate students who had taken the test in the past	Each was interviewed 3x (about 90 min. for each): Asked to reflect on test-taking experiences & how the experiences affected their lives.	Not controlled; test has been taken at some point in the past	Narrative inquiry: authors wrote narratives based on interviews, used thick description, "thematically analyzed " (p. 240) the narratives, and used member checking (asked participants to read and approve paper draft).	The test gave participants an opportunity to change their status by studying or working abroad (positive legitimizer), or it constrained their status by restricting the university courses they could take (negative gatekeeper).	Standardized testing requirements can be a financial, emotional, & temporal burden that in multiple ways disadvantage international students.

9. Hamid et al., 2019, *Language Testing in Asia*	IELTS	As reported earlier in Hamid & Hoang (2018), the study, explained in full here, included 430 prior IELTS test takers from 52 countries.	Responded to a survey; within the survey they were asked to comment on test issues and/or provide suggestions for test improvement.	Not controlled; Surveys were administered independent of actual test event	Only open-ended data analyzed; used "qualitative coding and content analysis:" responses coded at phrase, sentence, and paragraph level. Second coder coded all; agreed "over 98%" (no page #s).	Test takers stressed they needed much test-wiseness to do as well as expected. They also discussed the financial stress & emotional burdens the test put on them.	Investigating how tests like IELTS impact the test takers' lives is imperative for social justice & equality in language assessment.
10. Kang, Miao, & Hirschi, 2024, *RELC Journal*	IELTS	241 college-aged students from US, South Korea, & China who had taken the IELTS listening test in the last 12 months or who planned to take the IELTS test in the future	They provided acceptability ratings & comments regarding potential accents that could appear on the IELTS listening test.	Ratings & comments gathered right after listening to the simulated (potential) listening test accents	Exploratory factor analysis (EFA) used on the Likert-scale acceptability ratings. Test takers' qualitative comments presented to explain the numerical findings. [Method not identified.]	Regardless of geographic background, test takers preferred L1-English accents, that is, the accents they listened to while preparing for the test, & that they expected to hear at an English-medium university abroad.	Next may be to investigate the impact of strategies to align tests with test takers' experiences, or to investigate the impact of strategies that are put in place to sway test takers' views.

We observed the following three important overall trends in the research of the ten papers (see Table 1) that informed us on the design of our current study. First, some studies (Fox & Cheng, 2015; Kim, 2017) asked participants to provide their views of a test they took in the past *or* that they would take in the future (for which they were preparing), or they elicited views on more than one test (Cheng & DeLuca, 2011; Fox & Cheng, 2015; Hamid & Hoang, 2018); these approaches made it difficult for the authors to use the data to suggest concrete suggestions for revisions to any particular test. Only three of the ten studies (Elder et al., 2002; Kang, Miao, & Hirschi, 2024; Winke et al., 2018) had students actually take the test under investigation in a controlled environment and *then* had the students provide their perceptions within a short, regulated amount of time after the test was taken.

Second, we noted that Likert-scale questionnaires (Elder et al., 2002; Noubandegani, 2012), surveys (Fox & Cheng, 2015; Hamid & Hoang, 2018; Hamid et al., 2019) and one-on-one interviews with test takers (Sinclair et al., 2019; Winke et al., 2018) were prevalent in the research. One study included focus groups (Hamid & Hoang, 2018). However, those focus group data were neither analyzed nor part of the paper's results. Focus groups can be empowering to the focus group members and provide responses that reflect a society's views (Krueger & Casey, 2015). One modern goal of collecting test takers' views is to empower the test takers through a *participatory approach* (Little & Erickson, 2015); thus, more focus group research on test taker perceptions may be useful to fully understand the social consequences and local impact of proficiency tests, a goal Shohamy (2022) and Jin (2023) acknowledged as important.

Third and finally, we noticed that many of the studies used data analysis methodologies that were qualitatively underspecified (e.g., Hamid & Hoang, 2018; Kang, Miao, & Hirschi, 2024), or the researchers took a mainly quantitative approach to analyzing their perception data (Elder et al., 2002; Kang, Miao, & Hirschi, 2024; Kim, 2017; Noubandegani, 2012), which we believe may have short-changed the depth, richness, and nuances of what perception research should be able to offer. To be more fully specified, qualitative studies should include the study's theoretical framework, the researchers'

positionality within the study, and a description of how the methodology aligns with the framework (Braun & Clarke, 2024; Krueger & Casey, 2015; Terry et al., 2017). Richer and more descriptive methodology would provide more rigorous and nuanced qualitative results, such as the exemplary thick description of perception data by Sinclair et al. (2019), who provided narrative views of four test takers' perceptions after interviewing each participant three times about their recent ELPT experience.

In summary, Little and Erickson (2015), Jin (2023), Shohamy (2022), and O'Sullivan, (2012) have called upon test designers to collect and *apply* test takers' views within an iterative test design process. Doing so can empower the test takers and help improve the local, social impacts of tests. However, as reported by Hamid and Hoang (2018, p. 4), "we know little about how global English tests affect the lives and dreams of test takers," and thus research on test takers' views is needed, especially in locations where such research has not been conducted before. Thus, our research questions were:

1. What do West African, francophone learners of English report as their perceptions of a standardized, international English test after having taken it?
2. What are the test takers' recommendations for test design, administration, or score use?

Methodology

Study design

We designed a qualitative, focus-group-based study (Krueger & Casey, 2015) to answer our research questions. As described by Krueger and Casey, researchers conduct focus groups to better understand how people feel or think about an issue, idea, product, or service. Focus group participants should have characteristics in common that relate to the topic of the focus group. In our case, they were similar in age, at the same university, all learning English, and had taken the Duolingo English Test (DET) (see below) in November 2023 as part of this study. According to Krueger and Casey, each focus group should have about five to six participants and a trained moderator to guide the focus group, using a general template or script to launch discussions. Careful analysis of the

multiple discussions reveals patterns and trends that can be used to describe the groups' general opinions about the issue at hand.

Duolingo English Test

The DET is a one-hour, online English proficiency test that was launched in 2016. It was developed to make standardized English-language testing computer-adaptive, accessible to every test taker throughout the world, and affordable (Kostromitina & Wodzak, 2024). With the DET and other online English tests (such as the Test of English as a Foreign Language (TOEFL) Home Edition; ETS, 2025), test takers no longer need to travel to a proctored test center. Rather, they can take the test at home with a computer (laptop or desktop) with high quality audio, a webcam, a smartphone for recording the testing session (as a virtual proctor), and a high-speed internet connection. The computer must also be compatible with the test company's test-delivery application. Test takers can preview whether they have adequate equipment to take the test at home by following the testing company's equipment-check instructions online (see, for example, Duolingo English Test Help Desk, n.d.).

In particular, the DET evaluates "a test taker's ability to use linguistic skills that are required for speaking, writing, reading, and listening, as well as the integration of these skills" (Michalowski et al., 2024, p. 2). Scores are reported out as a total English-language proficiency score, and also as reading, writing, listening, and speaking subscores (Duolingo, 2024a). The scores are intended to be used by English-medium universities to estimate whether a test taker's English is proficient enough for admission (Park et al., 2023). DET scores are accepted by more than 4,000 universities within North America alone (Duolingo English Test, 2023a).

At the time of the study, the DET had a variety of question types, such as C-test (fill-in-the-blank) sentence- or passage-completion tasks, elicited imitation tasks, dictation tasks, and extended speaking and writing tasks, but some changes have been made since then; in particular, we were informed by an anonymous reviewer at the time of this paper's final review that the elicited imitation is being eliminated. The tasks that were present on the test at the time of our data collection can be viewed at Duolingo English

Test (2023b). How scores from the question types were assembled into the four subscores of speaking, writing, listening, and reading is detailed in Duolingo (2024). For example, the *speaking* score for the participants in this study was derived from a test taker's elicited imitation (read-aloud) tasks (wherein the test taker spoke a written sentence out loud) and the test taker's responses to five speaking prompts: one being a picture description task, and four being independent speaking tasks, with written and aural prompts. Alongside the overall score and subscores, the DET provides university admission offices and faculty members with open-ended samples of the test takers' speaking and writing for further evaluation and decision-making purposes through the university's secure DET dashboard (Duolingo, 2024a).

Test takers can further prepare by taking a free practice test that is available for delivery on mobile phones (Wodzak, 2023), although the real test requires, as mentioned above, a desktop or laptop computer plus a mobile phone to be used as a virtual proctor. The DET scores have been aligned to the Common European Framework of Reference (Council of Europe, 2020) and have been shown to correlate with scores from other commercially available tests (DET and IELTS at .78, DET and TOEFL at .77; Cardwell et al., 2023). Recent research (Isbell et al., 2024) has suggested that there is an appropriate extrapolation of DET speaking scores to real-life tasks that are part of an English-university student's academic-life experiences.

Study context

We conducted our research in Mali, Francophone West Africa, where French is the working language and where 13 local languages are recognized by Mali as official languages. At the university where we conducted this research, English is the medium of instruction for the students whose major is English. Concerning English testing in Mali, neither the central ministry of education nor the English department at the university have local proficiency or placement tests for the students. Instead of placement tests, the level of English instruction one receives is automatically based on the class level of the student. In Mali, exams are in paper format and teachers commonly use multiple-choice

tests because the classes are large (the minimum is 35 and the maximum is 80 or 100 students per class) and multiple-choice tests are easier to grade.

Additionally, the subfield of English language testing is still in its infancy in Francophone West Africa. Local policymakers have not yet decided whether to invest in this field so that tests in Francophone West Africa reflect local values and beliefs. However, English is currently enjoying a privileged status in Francophone West Africa because the central ministries of education alongside the African and Malagasy Council for Higher Education (see https://hauniversity.org/en/CAMES.shtml) have established English as a compulsory language starting from the elementary schools through to universities, regardless of students' majors. However, little investment has been made to substantially revise English curricula and assessment methods to reflect these transformations. Amid these contextual complexities, Malian English learners want to take a standardized, international test to know their proficiency levels for the following purposes: to prepare and apply for study-abroad programs (United States' Fulbright, British Council's Chevening, and so on), to immigrate to Inner Circle countries in search of a better life, to work for non-governmental organizations as interpreters or translators, to become English teachers, and to escape from unemployment after their studies (Diallo, 2023). These are their main English learning goals, and they explain why the participants of this study signed up to take a standardized, international test, specifically DET, for the first time as part of our research.

Researchers' positionality

It is necessary to describe our positionality because it helps elucidate potential biases we may have as researchers (Terry et al., 2017). We, the authors, are language testing and applied linguistics researchers and teacher-researchers whose objective is to investigate whether standardized tests are meaningful and fair for the intended test taker populations. We believe that if a test claims it is an international proficiency test, its scores should be meaningful for all intended populations, including those from Francophone West Africa. To this end, the first author (a Malian teacher-researcher and a visiting professor at the second and third authors' university during the 2022-2023 academic

year) discussed this research when it was at a proposal stage with testing company representatives at the Language Testing Research Colloquium (LTRC) in New York in June of 2023. At that time, we were interested in finding a testing company that would allow us to give their test for free to test takers in Mali for the purposes of this research. At the conference, DET researchers expressed interest and proposed a viable way for our participants to take the DET in Mali, West Africa. In fall 2023, Duolingo provided us, the authors, with 72 free coupons for our participants to take the DET. Duolingo also provided us funding to travel to the 2024 LTRC in Innsbruck, Austria, to present a preliminary draft of our study findings. When we signed the funding agreement with Duolingo through the second and third author's university, it was documented that we have freedom to publish our findings after a technical review of the manuscript at its pre-print or "original article" stage, that is, prior to submission to this journal's editors for publication consideration (to view the pre-print version, see Koné et al., 2024). No further technical reviews by Duolingo were required after submission of the original article.

As we wrote at the beginning of this article, we were personally interested in what Francophone West Africans would think of a standardized ELPT if they had the chance to take one given the first author's prior experience taking an ELPT in Francophone West Africa, and due to the second and third authors' current and prior work in teaching English as a second language (ESL), in training ESL teachers, and their work in standardized, English-language-test development research.

Participants

We invited 72 students to sign up for and take the DET. As a whole, 64 students (23 graduate and 41 undergraduate) out of the 72 who were initially invited successfully took the test and participated in one of 10 focus groups. The remaining eight students (2 graduate and 6 undergraduate) did not complete the DET because of technical difficulties (i.e., internet connectivity issues due to power outages); their three attempts expired before they succeeded in taking the entire test. The participants were all Malian, but three of them had two nationalities, with their second nationality being a neighboring or nearby country. Out of the 64, there were 14 female and 50 male participants. Their ages ranged

from 20 to 46 with 25.50 as their mean age. Our sample of 64 included undergraduate students who were pre-service English teachers (n = 41; 32 male, 9 female); graduate students who were teacher trainees (in apprenticeships or teaching provisionally) (n = 12; 2 female, 10 male); and students in translation studies (n = 11; 3 female, 8 male). All were studying for their undergraduate or graduate degrees in teaching English (see Table 2). They were multilingual students, speaking at least three or four languages, with French being the working language, English the first compulsory foreign language, and one or two local languages, including Bamanankan, the dominant language in Bamako, the capital of Mali. Other local languages the participants spoke were Bomu, Dogon, Fulfulde, Mamara, Maninkakan, Mossi, Khassonké, Songhoï, Soninké, and Syenara.

Table 2. Participants distribution per focus group

Participants	No.	Number of participants per focus group									
Focus Group		1	2	3	4	5	6	7	8	9	10
Grad. Student Teacher Trainees	12(2)	0	0	0	0	2(1)	5	4	1(1)	0	0
Grad. Students Trans. Studies	11(3)	3(1)	0	1(1)	1	2 (1)	0	0	0	3	2
Undergrad students:	41(9)	3(2)	6(1)	4(1)	6(1)	2	1	3(2)	5	5(1)	5(1)
Total	64(14)	6	6	5	7	6	6	7	6	8	7

Note. No. = number. Grad. = graduate. Trans. = translation. The number in parentheses after the number of participants is the number of females in the group.

Data collection instrument: focus group interviews

To prepare for focus group interviews, we developed a protocol and script (see the Appendix). We tried out the protocol with three faculty members – the first author, Kadidja Koné, and two of her English-language teaching colleagues – and with the second author Paula Winke as the moderator. This led to some small adjustments to the script.

In November 2023, our 64 participants took the research-based (non-proctored) DET in 10 groups of 6 to 8 students (see Table 2). Once a group finished taking the one-hour DET, we organized the group's focus group interview for later in the same day. Each focus group was led by two researchers, one serving as the main moderator (Kadidja Koné), and one serving as the moderator's aid and technology assistant (Paula Winke). Each session lasted 60 to 90 minutes and was recorded online through Zoom with the participants'

informed consent. The focus group interviews were done in English, although Kadidja Koné encouraged the participants to speak French or one of their local languages if they had difficulty expressing their ideas in English. Two or three days later, the students' unofficial scores (their DET overall score and their individual subscores) were sent to them.

Data analysis

Kadidja Koné and Paula Winke worked with a TESOL and applied linguistics graduate research assistant, Matt Gordon (the third author), to help with the data analysis and to triangulate results. Matt uploaded the focus group videos to MAXQDA version 24. We purchased 20 *MAXQDA Transcription* hours, and Matt used approximately 15 of them to automatically transcribe the videos within MAXQDA. Kadidja and Paula then met with Matt over Zoom weekly over a four-month period to do the coding. We applied a deductive, coded-thematic analysis (Terry et al., 2017) with initial and final codes emerging from the data as shown in Table 3. The four numbered codes in the "Final Codes" column are the main codes, while the two cross-reference codes are those that were overlain over (co-occurred across) the four main codes.

Table 3. *Initial and final themes*

Initial Codes	Final Codes
1. Contents in the test 2. Test components or features	1. Test features/areas
3. Technology 4. Test delivery	2. Technology
5. Emotional direction and intensity 6. Anxiety	Cross-reference 1: Emotional direction/ anxiety
7. Inclusivity 8. Relevancy	3. Inclusivity
9. Opportunities to participate in English learning	4. Professionalization
10. Recommendations from learners to Duolingo/Test designers 11. Suggestions 12. Super interesting comments	Cross-reference 2: Recommendations

The data were first segmented automatically by MAXQDA by voice, that is, each segment was defined as the connected speech of a single speaker. Matt Gordon manually verified each segment was attributed correctly, and any un-attributable turn or undecipherable speech segment was coded as participant 999--a dummy code for data not to be analyzed, which accounted for less than 1% of the recorded data. An example of 999 data would be when two or more speakers spoke at the same time, making individual speech unidentifiable. Identifying and segmenting 999 (un-usable) data is a normal part of data cleaning: Researchers can either delete such data prior to data analysis, or leave it in the dataset coded as unusable; we left it in as 999.

After data cleaning, the three authors discussed potential themes from the focus group sessions. Matt Gordon then entered those themes as codes in MAXQDA as shown in the left column in Table 3, and he worked to apply the themes to the data. Each week the researchers met up to discuss the application of the themes, the discrepancies, and how to solve them, thus basing the coding within a socialization (and triangulated) process. All three authors discussed code trimming, splitting, or consolidation. After a four-month iterative process (February through May 2024), the coding of the data concluded, with the final codes being in the right column of Table 3.

Findings and discussion

Deductive, coded thematic analysis of the qualitative focus group-based data indicated that the participants' discussions focused on four themes: (1) test features/areas, (2) technology, (3) inclusivity, and (4) professionalization. There were also two cross-references (that is, codes that were applied across the other codes): (1) emotional direction/anxiety and (2) recommendations on high-stakes English proficiency testing. We will present each theme with a summary of the findings, followed by a discussion of these results. For each theme, we will display one or two quotes from the participants to illustrate it, although the entire dataset was taken into account for the selection of the themes. In this study, we gave each participant an identification number preceded by S (undergraduate student) or T (graduate student teaching English as an apprentice).

Theme 1: Test features/areas

Our participants made 439 comments on the test's features and areas of evaluation. Their comments mainly focused on time (64 comments), picture description (91), the four skills (58) that they referred to as integrated skills, and accents (60). Regarding time, the participants commonly appreciated the one-hour length of the test that they believed was less stressful compared with other high-stakes tests, that may last three hours or more. However, some participants explained that the time allotted to complete certain activities was oftentimes short and they were obliged to rush, meaning that they might not score high in these parts, as opined by S03:

> You know, the time allocated to answer the question was really very short for a non-native speaker of the language. Because we need to think, we need to understand the questions and try to come back. So really this part is really tricky for me.

Also, the findings indicated that the participants believed that the DET measured all the language skills, from speaking, listening, reading to writing, which they perceived as good, as S03 highlighted:

> I find the test very interesting because, actually, I could see really the four skills of language learning in it. So the test, uh, we were, you know, uh, speaking parts, writings, listening, and, you know, the four skills were here.

These comments may be contextualized by knowing that these participants are most familiar with multiple-choice questions. Also, exams in Mali are in paper format; for most of the test takers in this study, this was their first online ELPT.

For the picture description activities, most of the participants expressed a belief that the time constraint made the task difficult, but that the task was still feasible, as shown in the comment of S42: "To think about an image and give your point, it is a writing form, not speaking. This one minute is too short in my view." Others, however, enjoyed the novelty in the picture description task, and noted that they would use such tasks in their current or future English language teaching.

Regarding accents, the participants stressed that it was easy for them to understand the speakers, even though they were mostly North American speakers of English, which is not the most common accent they encounter as learners. Test taker T59 explained: "Different accents during the test were very comprehensive, if you listen to them." Like the participants in Kang, Miao, and Hirschi (2024) (see Table 1), they connected intelligibility with the L1 accents they heard. Accordingly, there were some calls to diversify the accents and contents for the Francophone West African test-taking population. For example, some students wondered out loud if the test could be adapted to their geographical and learning context. Since they had to report their nationality in the background information before taking the test, and because the actual location of online testing is recorded (Duolingo, 2024b), they wondered if the content could be generated for them on the spot using that information, with more African accents, contexts, and images being selected to comprise their test.

Theme 2: Technology

The participants made 150 comments on technology. They commonly mentioned the difficulties they met during the registration process to take the test (28 comments), computer use (43 comments), and network connection (28 comments). We recorded that 80 percent of the students in our study did not have a personal laptop or a computer: These individuals borrowed laptops from us to take the test (we had five for them to use). Overall, the students did not have consistent, high-speed network connections while

taking the test, mostly due to city-wide, micro power outages (of from a couple of seconds to multiple minutes) that occurred throughout the day. These outages are part of normal life in Mali, reflecting the state of the electrical grid there, and they interfered with the participants' test-taking experiences.

Computer literacy was a major issue for the participants. While the majority were familiar with pen-and-paper testing, many had never used a computer or a laptop (or a keyboard and mouse either) before taking the DET. Test taker S27 asserted: "Even though I may have some ideas or arguments about the topic, I may have some problems about typing because some of us do not master computer science." Because universities in Mali do not use email and have no university email systems, most students did not have email accounts; they use mobile-phone-based apps for communication instead. Thus, part of the test preparation for us as researchers included helping students set up email accounts so that they could register with the DET and sign up to take their test. Most laptops in the region are second-hand and older machines; therefore, other technical difficulties surfaced. For example, for the registration the participants needed to show their identity card to the computer's camera, a task that was unexpected by the learners and difficult to do. We surmised that the difficulties in successfully scanning their identity-card photos may have been because the laptop cameras were not high resolution. For example, some students made many attempts before their identity was successfully checked and they proceeded to take the test. This situation caused stress and anxiety, as reported by the participants.

The network connection interruptions and power outages also negatively impacted the participants' test-taking experiences. Such issues caused them to retake the test (start over) whenever the speed of the connection was too low or when there was a micro power outage. It is worth noting that they were given three attempts to succeed. The researchers had rented a private power generator to supply the testing room with electricity,[1] but the time the generator took to get connected in case of a power outage was not fast enough for the connection to remain uninterrupted. These aspects negatively affected the participants emotionally and made them think that they would not score high, as argued by S20:

> You know, when, uh, having the test of Duolingo, anytime you have a problem with connection or power cut, it makes you start from scratch, you start doing again and again. Sometimes it can affect our emotions. It can affect, you know, uh, the your focus. You don't, uh, you know, when you have to do that again and again at the end, you will be quite, uh, discouraged about the test or something like this. So it may affect the test results.

What is interesting is that these negative situations did not decrease the participants' enthusiasm for taking the DET. The test takers provided overwhelmingly positive comments concerning the opportunity to take the test. They enthusiastically recommended that the DET become an application (mobile phone app) so that they could use their cell phones to take it instead of on computers. S46 suggested, "We don't have access to PCs [laptop]. And if the test could be done on a phone, that will be a lot much easier for us." Another advantage of using a mobile phone for ELPT is that electricity or internet-connection interruptions may not affect the testing if the test takers begin the exam with a fully charged phone and if they have a cellular-based internet pass that is sufficiently long for the duration of the test. Such conditions are possible and affordable in Francophone West Africa, and would circumvent the lack of availability of computers and the instability of the electrical grid.

In line with previous research (e.g., Fox & Cheng, 2015) that indicated that the international test takers' experiences were affected by variables such as familiarity with the computer keyboard, our findings also suggested technological issues, such as computer literacy, interfered with our participants' test-taking experiences. However, our results demonstrated that the test takers from Francophone West Africa and the test designers or testing companies are not on the same page technology-wise. It is not that the test takers are computer illiterate, it is rather that technology acumen and technology uses and needs are *different* in Francophone West Africa. The systems in Mali are much more mobile-based. Malians, for example, do not have the economic means to have access to the test designers' technological infrastructures. In short, the findings under the theme of technology imply that the field of language testing needs to leverage and promote the type of technology (i.e., cellular phones or any other technological infrastructures) used in Francophone West Africa to design tests that align with the current trends in the region. Further research would have to be conducted to understand how mobile-phone-based testing would impact test-score interpretation, much like how the field saw much research

on computer-based testing versus paper-based testing when computer-based testing first emerged. For example, would mobile-phone-based scores need to be reported out on a different scale from computer-based scores? Do mobile- and laptop-based testing have concurrent validity? When test scores are used for university-entrance examinations, is computer literacy (laptop or desktop computer-use knowledge) *supposed* to be part of the construct? Thus, would scores from mobile-phone testing distort score-meaning if there is not a separate, mobile-phone scale?

Meanwhile, other fields such as public policy, health sciences, medicine, and agricultural sciences are successfully using cellular phones for information sharing and infrastructure building in Africa (Rotondi et al., 2020). We do believe that adapting a test to the test takers' realities, including technological infrastructures, will not decrease the test's validity, but rather will add to it. It will align the test with the test takers' local practices and contexts, as advocated by O' Sullivan (2012), thus promoting equity which will reinforce learners' feeling of belonging to the international testing and World Englishes community. This positioning would help Francophone West Africa's test takers be able to take the test to invest in their English learning and to achieve their learning goals, with the test seen as a *legitimizer* and not a *gatekeeper*, as Sinclair et al. (2019) explained ELPTs can be.

Theme 3: Inclusivity

The analysis showed 135 comments on how the test made the participants feel included or excluded. Most of these comments were related to representation (26) and learning contexts (85). In terms of inclusion, certain participants reported that the test was accessible due to the diversity of the topics proposed to them for the speaking and picture description activities. As asserted by T63, the test is "inclusive and uh, since it really takes into account, uh, all language skills and, um, is what I really do appreciate, being a teacher or with, uh, students, they can, uh, find themselves in it [DET]." Many test takers noted that the test items were innovative and inspired them to engage in more online assessment and online language-skills practice. As noted by T64, "questions and structures about the test were really relevant to our level." S30 said they appreciated the

project "for giving us the opportunity to come across this kind of test… it's really, I can say, it plays a paramount role during our, I can say, educational process. It's not easy to come across these kinds of things for free."

However, some students opined that their region was not represented in the standardized test. For example, they explained that the test did not appear to have images or pictures based on their Francophone West African culture, and some tasks were a bit confusing because of the cultural differences. Particularly, T70 noted an image-description task happened to have included musical instruments that are not found in Mali, and explained "I'm sure I put nonsense there… because I couldn't understand these music instruments, and it's linked to culture." Likewise, images of Western playgrounds or sports were explained as difficult to describe, due to a lack of familiarity. S37's comment summarized some of the sentiments of the test takers regarding inclusivity:

> I think it will be very better if they include images and topics, uh, from every continent, Africa, Europe, America, Asia. If they include images and topics from everywhere, I think it will make it more perfect because everyone will see, see himself in it. If an African, uh, image or topic, a topic about Africa, I will say that they take care about Africa and I can see myself in [it], I, I will feel concerned about the test.

The teacher trainees particularly mentioned that the learning contexts may be a challenge for students in Francophone West Africa. In educational contexts in Mali, students are tracked into science or letters from the tenth grade (approximately age 15). The students whose field is science take courses that mainly focus on mathematics, biology, physics, and chemistry, whereas those whose major is letters have French literature, English, and another foreign language as their main subjects. From the teachers' point of view, it can be challenging for their students to score high when the activities target a specific domain that is out of their field of study.

Overall, under the theme of inclusivity, the participants further recommended that the test designers include British accents or English spoken by local West Africans to make the test less American-centered. In our study, learners were not taking the test to study abroad, but rather to assess their own English on an international scale. Their comments pointed out that they believed that a test with only North American accents would *not* be

the best estimator of their proficiency for this purpose (for proficiency testing and/or for diagnostic purposes). Our test takers' opinions on accents in the DET aligned with another speech perception study by Kang, Yan, et al. (2024) who found listeners understood the English of speakers who shared their (the listeners') L1 better than they did the English of speakers who did not share their L1. The test takers in our study have mostly listened to British-accented English and local West African speakers of English when learning English in Mali, although they have had exposure to North-American-accented English as well through popular culture. Thus, some of the test takers had a preference to have their English assessed through those three accent varieties, and not just through North American accents, to better represent their English abilities and to align with how English is taught and learned in Mali.

Importantly, the participants of the present study expressed strong opinions that they enjoyed taking the test because they wanted to understand where they were on an international language proficiency scale. Knowing about their level on an international framework, such as the Common European Framework of Reference for Languages (CEFR), develops their agency and helps them construct or reconstruct their identity (as corroborated by Little & Erickson, 2015). Many participants in this study felt the ELPT opened doors and aligned with their more localized and contextualized goals, an inclusivity theme aspect that overlaps with Theme 4 on professionalization, which we describe next.

Theme 4: Professionalization

The participants reported that the DET offered them the opportunity to develop professionally–that is, to plan on how to improve their English language skills (undergraduate students) and also how to choose the contents of their English courses (graduate students)–and become a member of the international test taker community. Accordingly, there were 47 comments on self-improvement and 64 comments on gratitude toward the researchers and the testing company for having provided them with the opportunity to take the test. Regarding self-improvement, the participants explained that the test successfully assessed their English skills, which they believed boosted their

self-confidence and pride as English learners. For example, S24 stated: "So, uh, I said this test evaluates our ability [skills]. To read, uh, speak, listen, and write in English. Yeah. After doing so, we, uh, got self-confidence. Yeah. We, uh, and I can say I'm proud of myself." S04, additionally, pointed out that the test offered opportunities to know areas in which they should work to improve their English skills, as a diagnostic tool: "The way it's structured is good. You will know, uh, what you are good at and what you have to still work on. And, yeah, I think the test is great."

Most of the test takers had not taken any international standardized tests before the present study. Therefore, expressing their gratitude may have been a way to show that they appreciated taking the test and that they wanted to pay tribute to the persons who provided them with such "a golden opportunity" to take an ELPT, as revealed by S26:

> I would firstly thank them [researchers and testing company] for what they have done because it's a really interesting, uh when you master this, it can give you an overview, an idea about all of the other tests, online tests.

Some participants were grateful because the test allowed them to carry out some interactive speaking activities in an environment that normally does not provide extensive opportunities for oral practice:

> I think that this uh test is a big opportunity for us because here we are in a French speaking country. It is not easy to learn English. And I think that those kinds of activities are an opportunity for us to practice our English more. (S54)

Participants within the current study suggested that the testing company and researchers continue involving them in international testing so that they could understand how to use the test scores for diagnostic purposes – for themselves, and for their students – and to envision how to embed assessments within their curriculum. Other participants noted they would like to take the test again, but to do so, they must be able to pay appropriate-level test fees using Western Union or Moneygram, which are international payment systems that are available in their contexts. Within Francophone West Africa, credit cards issued by local banks cannot be used to make international purchases (like a test payment), and very few students and teachers have an international bank card. The gross domestic product per capita in Mali is, according to the World Bank, currently about 833

USD, which is much more than students earn per year. Thus, students noted that a fee of 5 USD may make the test more accessible and usable for local purposes.

Limitations and future directions

Our study has limitations that we would like to discuss in hopes that they may inform future researchers. First, although we are the first, to our knowledge, to conduct standardized ELPT perception research in Francophone West Africa (see Table 1), our research involved a convenience sample, which limits the extrapolation of our findings. Second, all test takers in our study were current or future English teachers or professionals who will be working in English-language translation. As pointed out by a reviewer, our results may have been different or differently informative if we had included West African test takers who would be interested in taking the test to study science or engineering abroad, for example, at an institution in North America that accepts DET scores for international student admissions. Thus, a future ELPT perception study with Francophone West Africans could take a methodological approach similar to that of Sinclair et al. (2019): they interviewed four individuals already studying abroad who had taken an ELPT as part of their admissions process. Sinclair et al. asked the students what they thought about the English test that they had taken and how it had impacted them. Such a narrative study could be replicated with Francophone West Africans who took the test specifically for college or university entrance abroad.

Despite these limitations, we believe the results are useful for ensuring international, standardized, English-language tests in Francophone West Africa are *positive legitimizers* (Sinclair et al., 2019), meaning that scores from them encourage pathways forward, with scores serving as meaningful *proficiency indicators,* charting the test takers' World-English development on an international scale.

Conclusion

Until our study, no research had examined the beliefs and experiences of test takers from Francophone West Africa taking potentially high-stakes English-language tests.

However, it has been reported that this region has the world's youngest population with a high rate of population growth (Walsh, 2023). Additionally, "[t]he proportion of young Africans attending university has risen nine-fold since 1970" (McDermott, 2025), which may foreground growth in the number of African university students' interest in acquiring English as a World Language to an advanced or superior level. Thus, we assume this area of the world could be an extremely large and influential body of English-language test takers.

We had two aims that we converted to research questions: (1) to gather the Francophone West African test takers' views regarding the Duolingo English Test (DET) as an international, standardized English test (through focus groups administered shortly after the test takers have taken the test), and (2) to gain insights on how the test could potentially be revised in light of the test takers' views. The findings indicated that the Francophone West African test takers believed that the DET offered them an opportunity to evaluate their proficiency on an international proficiency scale. Also, taking the DET helped them to reconstruct their identity and position themselves as members of the international test taker community, especially since most of them had never taken such a test before, mainly because of its expense and the technological resources needed to take the test online. In sum, having taken the DET empowered (or *legitimized*; Sinclair et al., 2019) the West African learners of English to improve their skills in English even more. They felt like they belonged to the international community by being given the opportunity to take a standardized test, which is normally out of their financial and technological reach, but also they noted that the test content lacked West African contexts and that they, as West Africans (and their associated experiences and trajectory in learning English), were not well represented in the test material itself, something that they hoped might change in the future.

On the other hand, the findings indicated that the participants were confronted with technological difficulties (low internet connection due to power outage, computer use, etc.) while registering to take the test and while taking it. Nonetheless, these negative technological conditions that triggered stressful, emotional feelings did not decrease their enthusiasm and recognition that taking the standardized test was an opportunity: It was

an international test that integrated the four skills with a variety of activities related to vocabulary and picture descriptions. These items, although not aligned with the participants' curriculum, were sources of motivation to retake the test if offered another "golden opportunity" (S026), and to try out the tasks with their current or future students. The future teacher participants considered the test as a professional opportunity that they could use to inform their teaching, demonstrating standardized testing's strong washback power.

Concerning the second aim, the West Francophone African test takers suggested that the test designers and testing companies consider making under-one-hour, mobile-phone-based, lower-stakes tests that would report scores out on an international language proficiency scale (like CEFR or ACTFL) for Africa. Also, lower-quality/-resolution phone cameras and microphones would need to be allowed, as well as lenient e-proctoring/security. Others (e.g., Nguimkeu & Okou, 2021; Rotondi et al., 2020) have shown that international companies investing in mobile-phone infrastructure in West Africa are having success; thus, we believe testing companies can too. In the same vein, testing companies need to invest in customized (e.g., African-centric) content to make the standardized proficiency tests more relevant to African test takers. Francophone West African English language learners should be able to "see themselves" in the test, as opined by participant S37 in the results section.

Shohamy (2022) noted that investigating test takers' insights can or should involve monitoring test-score uses, and we found, through our research, that ELPT test-score uses other than what the testing companies may directly advise are prevalent: Our participants were not taking the test to study abroad (the primary, *on-label use* of this ELPT's scores), but rather to use the scores as a type of diagnostic. Based on the test takers' comments, we believe such local uses align with the broader definition of this ELPT as an *English language proficiency test*, and this diagnostic use could be prevalent in the region going forward. But for that use to take hold, students asked for the test fee to be as low as five dollars (for Francophone West Africa). Based on the students' contexts and insights, if implemented, such a fee would promote equity for the Francophone West African test takers and better welcome them to the subfield of language testing and assessment.

Certain testing companies such as Duolingo have recently established fee waivers for low-income test takers applying to the universities of inner circle countries (USA, Canada, United Kingdom) (Duolingo English Test, n.d.), but this solution may exclude the majority of the Francophone West African students who would like to take the test as an international measure of their proficiency and/or for diagnostic purposes: for the waiver, the university must ask for it on behalf of the applicant. Overall, as attested to by the participants, the current test fees are exorbitant for the majority of the Malian participants, which is problematic as tests should be financially accessible in order to be fair (Kunnan, 2008). However, it is worth mentioning that the DET is one of the least expensive and most accessible ELPTs on the market. Our participants were excited about taking it and could see retaking it as their English improves. For their continued participation, we wondered if testing companies could offer sliding-scale fees, much like how publishers and professional organizations do: for example, Cambridge University Press and the American Association for Applied Linguistics (AAAL) offer free open-access publishing and free membership, respectively, to individuals from World-Bank identified, low-income countries (see AAAL, n.d.; Grove, 2023). Can testing companies do the same? If fees must be kept, online payment infrastructures in Francophone West Africa such as Western Union, Moneygram, and Ria Money Transfer would need to be options for test registration because international credit cards are generally not available in this region. Such infrastructure investment by companies may seem daunting at the outset, but would pay off in the long run, as attested by the economist Ayenew (2022), who studied foreign investments over time in Africa. He wrote that short-term gains from these investments within Africa may be insignificant, but they pay off over time.

We end our study by asking international testing companies to persist in supporting Francophone West African researchers and testers by continuing to involve them in the global research on language testing. Companies can do this by supporting research on language testing (as our study was supported) and by funding conferences on language assessment issues in Africa (British Council, 2024; ILTA/Duolingo, 2023). Future research could examine how the use of new technologies in large-scale, high-stakes testing will affect Francophone West African test takers' identity (re)construction or agency and could investigate the use of standardized, international language proficiency

frameworks, such as the CEFR, in the region. Questions abound, but we know from this research that the theorists (Jin, 2024; Kane, 2013; O'Sullivan, 2012; Shohamy, 2022) are right: test takers' views are essential to answering the questions and for ensuring that test score interpretations are meaningful and fair.

Note

Normally the DET is taken by test takers in a secure room with no one else in the room. However, for this research, Duolingo allowed our test takers to take the test in a semi-secured, proctored environment, which was needed because most students did not have a computer nor a laptop. The first author set up a temporary computer room with five laptops in one of the university president's conference rooms, where students, in groups, could take the DET, and then later participate in a focus group.

Acknowledgements

We would like to thank Drs. Binta Koita and Fatoumata Keita at the *Université des Lettres et des Sciences Humaines de Bamako* for contributing to this study and for participating in the pilot focus group with us.

Author disclosures

This paper was sponsored with funding by Michigan State University's Office of International Studies and Program's Alliance for African Partnership (AAP), the University of Bamako's Office of the President, Michigan State University's College of Arts and Letters, and by a grant from Duolingo (Michigan State University account number RG101905). Duolingo provided the first two authors travel funding to present this research at the Language Testing Research Colloquium (LTRC) in Innsbruck Austria in July of 2024. Author contributor (CRediT) roles: Kadidja Koné: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualization, writing – original draft, writing – review & editing. Paula Winke: conceptualization, data curation, formal

analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualization, writing – original draft, writing – review & editing. Matt Gordon: data curation, formal analysis, methodology, project administration, validation, visualization, writing – original draft.

ORCID iDs

Kadidja Koné https://orcid.org/0000-0003-4209-9195

Paula Winke https://orcid.org/0000-0002-8169-650X

Matt Gordon https://orcid.org/0009-0001-2858-3995

References

AAAL. (n.d.). *Join AAAL!* American Association of Applied Linguistics. https://www.aaal.org/membership-information

Ayenew, B. B. (2022). The effect of foreign direct investment on the economic growth of African countries: An empirical approach. *Cogent Economics & Finance*, *10*: 2038862. https://doi.org/10.1080/23322039.2022.2038862

British Council. (2024). *New directions Africa 2024.* https://www.britishcouncil.co.ke/programmes/new-directions--africa-2024

Braun, V., & Clarke, V. (2024). Reporting guidelines for qualitative research: A values-based approach. *Qualitative Research in Psychology*, 22(2), 399-438. https://doi.org/10.1080/14780887.2024.2382244

Cardwell, R., Nydick, S., & Lockwood, J. R. (2023). *Considering inter-test relationships in high-stakes admissions testing: The case of English.* https://d23cwzsbkjbm45.cloudfront.net/media/resources/alte_23_cardwell.pdf

Cheng, L., & DeLuca, C. (2011). Voices from Test-takers: Further evidence for language assessment validation and use. *Educational Assessment*, *16*(2), 104–122. https://doi.org/10.1080/10627197.2011.584042

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. https://www.coe.int/en/web/common-european-framework-reference-languages

Diallo, R. (2023). Critical pedagogy and postmethod in West Francophone Africa: Possibilities and practical applications: The case of Mali. In G. Park, S. Bogdan, M. Rosa, & J. M. Navarro (Eds.), *Classroom pedagogy in the language and writing classroom: Strategies, examples, activities from teacher scholars*, (pp.119–134). Routledge. https://doi.org/10.4324/9781003357001-11

Duolingo. (2024a). *From 2020 to 2024. A look at the evolution of the DET*. https://duolingo-testcenter.s3.amazonaws.com/media/resources/Evolution-of-DET-2024.pdf

Duolingo. (2024b). *Privacy policy*. https://preview.duolingo.com/privacy

Duolingo English Test (2023a). *Accepted by thousands of institutions worldwide*. https://englishtest.duolingo.com/institutions

Duolingo English Test (2023b). *Duolingo English Test walkthrough 2023: Test overview with all question types*. https://youtu.be/avhEAqbccus?si=IQw3nNzyARMYC0RS

Duolingo English Test. (n.d.). *Mission in action: Apply for fee waivers*. https://englishtest.duolingo.com/institutions/access

Duolingo English Test Help Desk. (n.d.). *Duolingo English Test rules and requirements*. https://testcenter.zendesk.com/hc/en-us/sections/360001814091-Duolingo-English-Test-Rules-and-Requirements

Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: what does the test-taker have to offer? *Language Testing, 19*(4), 347–368. https://doi.org/10.1191/0265532202lt235oa

ETS. (2025). *TOEFL iBT testing options*. https://www.ets.org/toefl/test-takers/ibt/about/testing-options.html

Fox, J., & Cheng, L. (2015). Walk a mile in my shoes: Stakeholder accounts of test taking experience with a computer-administered test. *TESL Canada Journal*, *32*(9), 65–86. https://doi.org/10.18806/tesl.v32i0.1218

Grove, J. (2023, April 14). *Cambridge University Press drops fees for low-income countries*. Times Higher Education. https://www.timeshighereducation.com/news/cambridge-university-press-drops-fees-low-income-countries

Hamid, M. O., Hardy, I., & Reyes, V. (2019). Test-takers' perspectives on a global test of English: Questions of fairness, justice and validity. *Language Testing in Asia*, *9*, article 16. https://doi.org/10.1186/s40468-019-0092-9

Hamid, M. O., & Hoang, N. T. H. (2018). Humanising language testing. *TESL-EJ, 22*(1), 1-20. https://tesl-ej.org/wordpress/issues/volume22/ej85/ej85a5/

ILTA/ Duolingo. (2023, November 9). *Collaboration and outreach grant workshop: Empowering Francophone African women teachers to use performance-based assessment in their English language programs.* Université des Lettres et des Sciences Humaines de Bamako (ULSHB), Bamako, Mali.

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

Jin, Y. (2023). Test-taker insights for language assessment policies and practices. *Language Testing, 40*(1), 193-203. https://doi.org/10.1177/02655322221117136

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

Kang, O., Miao, Y., & Hirschi, K. (2024). Test-takers' perceptions of the inclusion of different English accent varieties in the IELTS listening tests. *RELC Journal*. Online advance publication. https://doi.org/10.1177/00336882241231310

Kang, O., Yan, X., Kostromitina, M., Thomson, R., & Isaacs, T. (2024). Fairness of using different English accents: The effect of shared L1s in listening tasks of the

Duolingo English test. *Language Testing, 41*(2), 263-289. https://doi.org/10.1177/02655322231179134

Kim, E.-Y. J. (2017). The TOEFL iBT writing: Korean students' perceptions of the TOEFL iBT writing test. *Assessing Writing, 33*, 1-11. https://doi.org/10.1016/j.asw.2017.02.001

Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). SAGE.

Koné, K., Winke, P., & Gordon, M. (2024). *"We would like to see ourselves in the test:" The experiences of francophone African English learners in high-stakes English proficiency testing*. Open Science Framework Preprints. https://doi.org/10.31219/osf.io/tsbf5

Kostromitina, M., & Wodzak, S. (2024, February 7). The evolution of the Duolingo English Test. *Test Center*. https://blog.englishtest.duolingo.com/duoling-english-test-different-versions/

Kunnan, A. (2008). Towards a model of test evaluation: Using the test fairness and test context frameworks. In L. Taylor & C. J. Weir (Eds.) *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity* (pp. 229–251). Cambridge University Press.

Little, D., & Erickson, G. (2015). Learner identity, learner agency, and the assessment of language proficiency: Some reflections prompted by the Common European Framework of Reference for Languages. *Annual Review of Applied Linguistics*, *35*, 120–139. https://doi.org/10.1017/S0267190514000300

McDermott, J. (2025, January 6). *Like the jaws of the crocodile… The economic gap between Africa and the rest of the world is growing*. The Economist. https://www.economist.com/special-report/2025/01/06/the-economic-gap-between-africa-and-the-rest-of-the-world-is-growing

Nguimkeu, P., & Okou, C. (2021). Leveraging digital technologies to boost productivity in the informal sector in Africa. *Review of Policy Research*, *38*(6). 707–731. https://doi.org/10.1111/ropr.12441

Michalowski, A., Cardwell, R., Nydick, S., Naismith, B. (2024). Duolingo English Test: Demographic and score properties of test takers. *Duolingo Research Report*. https://englishtest.duolingo.com/research

Noubandegani, P. A. (2012). Students' perceptions of computerized TOEFL test. *Language Testing in Asia*, *2*, Article 73., https://doi.org/10.1186/2229-0443-2-2-73

O'Sullivan, B. (2012). A brief history of language testing. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language assessment* (pp. 9–19). Cambridge University Press.

Park, Y., Cardwell, R., Goodwin, S., Naismith, B., Laflair, G. T., Lo, K.-L., & Yancey, K. P. (2023). *Assessing speaking on the Duolingo English Test*. https://englishtest.duolingo.com/research

Rambiritch, A. (2014). The accessibility of the test of academic literacy for postgraduate students (TALPS): student perceptions. *Per Linguam, 30*(1), 18–27. https://doi.org/10.5785/30-1-214

Rambiritch, A. (2015). Accountability issues in testing academic literacy: the case of the Test of Academic Literacy for Postgraduate Students (TALPS). *Perspectives in Education*, *33*(1), 26–44. https://doi.org/10.38140/pie.v33i1.1894

Rotondi, V., Kashyap, R., Pesando, L. M., Spinelli, S., & Billari, F. C. (2020). Leveraging mobile phones to attain sustainable development. *PNAS*, *117*(24), 13413–13420. https://doi.org/10.1073/pnas.1909326117

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*, *36*(2), 236–252. https://doi.org/10.1080/15434303.2019.1628238

Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. In C. Willig & W. Stainton-Rogers (Eds.), *The SAGE handbook of qualitative research in*

psychology (2nd ed.) (pp. 17–37), SAGE. https://doi.org/10.4135/9781526405555

Van der Walt, J. L., & Steyn, H. S. (2007). Pragmatic validation of a test of academic literacy at tertiary level. *Ensovoort*, *11*(2), 138–153. http://hdl.handle.net/10394/2841

Walsh, D. (2023, October 28). The world is becoming more African: Part one of the series on how the youth boom is changing the continent, and beyond. *The New York Times*. https://www.nytimes.com/interactive/2023/10/28/world/africa/africa-youth-population.html

Winke, P., Lee, S., Ahn, J.I., Choi, I., Cui, Y., & Yoon, H.-J. (2018). The cognitive validity of child English language tests: What young language learners and their native-speaking peers can reveal. *TESOL Quarterly, 52*(2), 274-303. https://doi.org/10.1002/tesq.396

Wodzak, S. (2023). *Three ways we improved the DET practice test.* Duolingo Test Center Blog. https://blog.englishtest.duolingo.com/3-ways-we-improved-the-det-practice-test

Appendix

Focus Group Session Protocol and Script

Part 1: 40 minutes

1. What did you like in the DET?
2. What did you not like in the DET?
3. Did anything surprise you in the DET?
4. What did you think of the various item types?
5. What did you think of the images?
6. Did you have any problems with the directions?
7. Do you feel that the test assessed your English?
8. Were you prepared to take the DET?
9. Are the DET scores meaningful to you?
10. If you could talk to the designers of the test, what would you want to talk to them about?

Part 2: 20 minutes

Next, we show you a series of 10 phrases one by one, each on a large card. Tell us specifically about the individual phrase, and what you think about the phrase in relation to the DET. We give you up to two minutes to discuss the phrase as a group after you are shown the phrase.

Getting set up to take the test (sound-check, etc.)

1. Test directions
2. Test length
3. Test availability in Africa (outside of this research project)
4. The audio files
5. The visuals on the test
6. The test scores and what they mean
7. The various item types (fill in the blank, speaking, writing)
8. The topics on the test (what you had to read, write, and speak about)
9. Diversity of people or ideas in the test images, and in the content of the test