CAMBRIDGE

Cambridge Papers in English Language Education

Remotely delivered language assessment: what makes it fit for purpose?

Hye-won Lee, Daniel R Isbell and Benjamin Kremmel

What will we cover in this paper?

Remote language assessment has grown from a niche alternative to a mainstream option because it offers greater convenience, flexibility, and accessibility; but its use must be carefully balanced against challenges such as security, bias, privacy, digital inequity, and the need for strong evidence that it is fit for purpose. This is not an argument against remote testing, but one for a more rigorous assessment of how, when, and in which contexts it can be used with confidence. With remote language testing also playing an increasingly important role in decisions such as those related to higher education admission or immigration, providers have a responsibility to ensure that convenience does not come at the expense of confidence in results.

While Cambridge English continues to examine the opportunities and risks of remotely delivered assessment, it has set out six foundational principles for best practice to support a more evidence-based approach to remote language testing.

Our six guiding principles:

1 Testing what matters

While test design is a fundamental consideration for any assessment, it becomes even more critical in remote settings. Practicality needs to be balanced against the need for a valid and meaningful evaluation of language proficiency. The test design and the tasks included must be carefully aligned with the specific language skills being evaluated and the overall purpose of the assessment.

2 Rigorous test security

Remote language tests need to place even greater emphasis on maintaining rigorous test security as they are more vulnerable to exam malpractice compared to in-centre tests. Providers need robust identity checks, effective monitoring, and clear protections against malpractice, alongside careful consideration for the appropriate level of remote proctoring, so that the results from remotely delivered tests can be trusted.

3 Standardised test conditions

Remotely delivered assessment sessions should follow consistent standards to ensure fairness for all test-takers. Since certain administrative aspects in remote settings fall outside the test provider's direct control, specific measures, such as prioritising user-friendly platforms, must be implemented to maintain consistency across sessions.

4 Maintain human involvement

Technology can support delivery and monitoring, but it cannot replace human judgement altogether. Appropriate human guidance, support, and oversight must remain integral to remotely delivered assessments.

5 Fair and inclusive participation

Remote testing can improve accessibility, but only if all candidates have a fair opportunity to demonstrate their competence. Remote delivery assessments need to incorporate differences in equipment, connectivity, environment and digital confidence to accommodate unique test-taker circumstances whenever possible.

6 Comparability across modes

When remote assessments are presented as an equivalent alternative to in-centre options, it is essential to validate and monitor the comparability of test quality and outcomes to ensure that scores from different delivery modes can be used interchangeably. It is also crucial that providers continuously assess the ongoing comparability of different delivery modes.

To maintain the integrity of remotely delivered language assessments, the principles outlined in this paper should be applied at every stage of testing practices. Providers, policymakers and institutions should also continue to take an evidence-based approach to ensure remote testing is secure, fair and fit for purpose.

Contents

Introduction

Historically, language assessments have been primarily administered in a paper-based format in test centres. Advancements in technology in the last 15 years have led to the evolution of remotely delivered test formats in language assessment, which offers improved administrative efficiency and on-demand assessment.

While remotely delivered language assessments have been available for several decades, starting with telephone-delivered speaking assessments in the 1990s, the Covid-19 pandemic accelerated the transformation of test delivery. Since then, remote testing (also known as 'online' or 'at home' testing) has evolved into a standard delivery option for assessments alongside in-centre delivery.

Given the growing use of remote assessment in high-stakes contexts and the polarised global trends in favour and against this assessment mode (e.g., Association of Test Publishers & National College Testing Association [ATP & NCTA], 2024; Ofqual, 2024), this is an opportune time to take stock of current practices and to discuss key considerations for best practice in remotely delivered language assessments. In this paper, we propose a number of foundational principles for best practice in remote assessment and illustrate their application using examples from Cambridge English tests. Our aim is to provide a balanced, critical perspective to contribute to debates around remote assessment and provide conceptual and practical insights. We aim to offer considerations, insights, and questions as a foundation for making informed decisions about the policy and practice of remote assessment.

Ultimately, the advantages of remote assessment need to be considered alongside its limitations. The choice of remote vs in-centre assessment has to be based on **sound considerations about the assessment context and purpose**, and informed by **robust evidence that remote assessment is fit for the context of use**.

Benefits and challenges of remotely delivered language assessment

There are a number of key advantages and challenges associated with remote assessment. At times, benefits coexist alongside constraints, reinforcing the need for nuanced considerations when making decisions about the advantages and downsides of using remote assessment.

Key advantages of remote assessment include:

- **Convenience and flexibility.** Remote testing eliminates the need for travel to test centres, which can be a burden for some test-takers, particularly in remote areas. Taking the test in a familiar environment may also help reduce anxiety often associated with unfamiliar testing settings. Additionally, remote testing can offer greater scheduling flexibility, allowing test-takers to choose test dates that can better suit their needs. For speaking examiners, remote delivery offers flexible working arrangements.
- **Increased accessibility and fairness.** By removing geographical barriers, remote assessment can extend availability to a broader population of test-takers. However, ensuring equitable access requires careful consideration of whether remote delivery enables fair, inclusive, and meaningful participation across diverse test-taker groups. Beyond location, remote delivery can also support other dimensions of access, including greater flexibility in test frequency and scheduling, potential reductions in indirect costs (e.g., travel), and improved access for individuals with disabilities or circumstances that hinder their ability to attend in-centre testing. (Although it is important to note that some accommodations, such as the provision of support personnel like scribes, can only be offered in an in-centre setting.) These access benefits exist alongside important limitations, which are considered elsewhere in the paper.
- **Preserving advantages without compromising integrity.** In the case of remote speaking assessments conducted via video calls, human/human interaction can be maintained, despite the physical distance between test-takers, examiners and administrators. This mode of delivery enables tests to assess communicative competence and ensures that the core elements of what is being assessed (the test construct), i.e., interactional competence, remain uncompromised by the remote format, while still preserving all the other benefits of remote testing.

Remotely delivered language assessments also present important challenges that require careful consideration to maximise benefits while minimising potential drawbacks:

- **Test security** is a central concern, particularly in high-stakes contexts. The remote nature of these assessments reduces the level of direct control and monitoring compared to centre-based testing, increasing the potential risk of security breaches. While centre-based testing is not immune to breaches, remote delivery can present its own distinct challenges. **Reliable detection of malpractice** in this context is essential to maintaining strong cybersecurity standards. This includes the effectiveness of AI-based monitoring tools, the training of remote proctors to ensure consistent and accurate identification of misconduct, robust ID checks, a secure lockdown browser, as well as post-test detection techniques such as statistical analysis of response patterns, anomaly detection, and biometric approaches (e.g., face and voice recognition) for imposter identification. Together, these measures form a layered approach to safeguarding test integrity.
- **Bias** is another critical concern in the context of remote proctoring. Some proctoring tools, particularly those that rely on AI, have historically raised concerns about racial or ethnic bias depending on how they are designed and trained (e.g., Burgess et al., 2022; Yoder-Himes et al., 2022). Although recent technological advances have improved the performance, robustness, and fairness of many commercial systems, these concerns have not been fully eliminated. If the datasets used to train such remote proctoring tools underrepresent certain demographic groups (i.e., groups that are insufficiently reflected in the training data), the systems may still perform less accurately for individuals from those groups and misinterpret normal behaviour as suspicious. Here, "underrepresent" refers to demographic groups that are insufficiently reflected in the training data, rather than any particular segment of the test-taking population. As a result, these individuals may be disproportionately flagged for potential cheating. Such bias can undermine the validity and perceived fairness of the assessment process and highlights the need for ongoing auditing, transparency, and evaluation as the technology continues to evolve.

- **Training remote proctors** can also present a challenge, particularly when the task is outsourced to third-party providers. Effective training is essential to ensure that proctors can reliably detect malpractice and provide the necessary support to help test-takers maintain fair and consistent testing conditions. However, training alone is not sufficient. Ongoing quality assurance and monitoring, such as periodic audits, review of proctoring decisions, and the use of "stealth tests" to assess proctor vigilance, are also crucial to maintaining consistent standards. In addition, robust vetting and selection of proctors can help ensure that those entrusted with overseeing high-stakes assessments have the appropriate skills and judgement. In the context of language assessment, training needs to additionally prepare proctors to work with test-takers with varying levels of language proficiency.
- **Privacy risks** may arise, as remote proctoring often involves collecting sensitive personal data, such as identification documents and video recordings of private spaces (e.g., Terpstra et al., 2023). Without stringent safeguards, this can lead to a risk of privacy violations or misuse of data. These challenges are further amplified in a global testing context, where differing national data protection laws and regulatory expectations can complicate the implementation of consistent privacy standards. These risks are especially serious for minors. Without clear privacy protections and active parental involvement, remote testing can expose minors to long-term data vulnerabilities, profiling, and surveillance.
- **Access to suitable technology and digital literacy** may introduce a degree of inequity, too. Remote assessments often require a certain level of digital literacy as well as access to reliable technology. This can disadvantage test-takers who lack the necessary resources or skills to fully benefit from this mode of delivery. Moreover, not everyone has access to a quiet, private space that meets the environmental requirements set by the test provider.

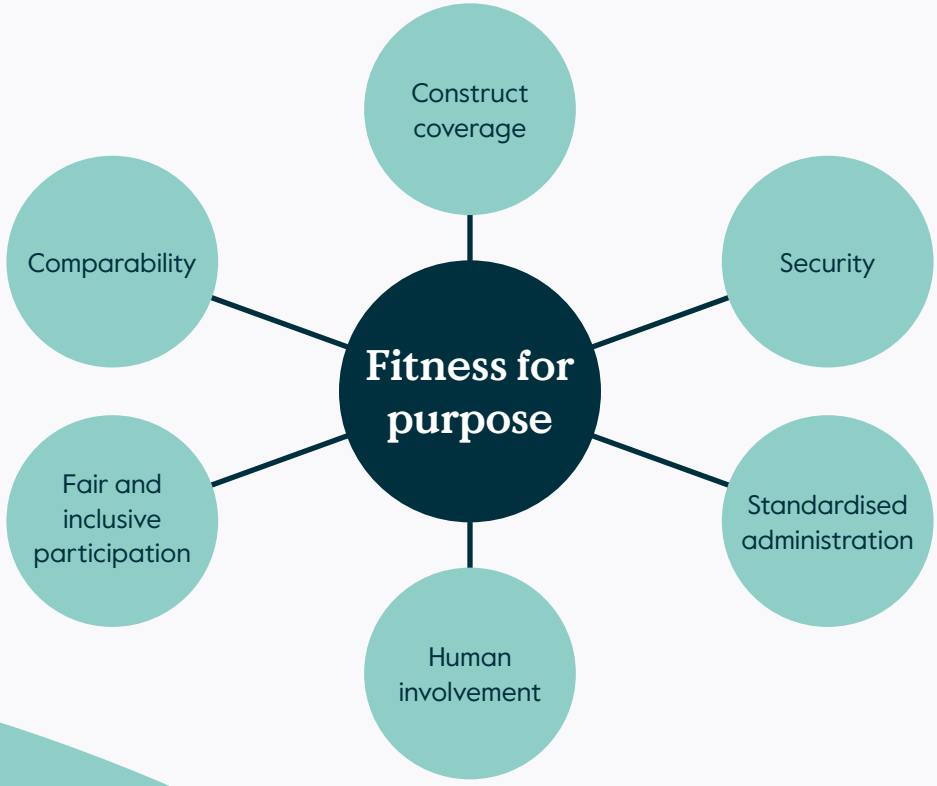
Considering the complex interplay of benefits and limitations associated with remote assessment, it is important that when remote assessment is chosen as the primary test delivery mode, there is a solid case for **why remote assessment is the optimal one to use and how the limitations and associated risks to the integrity of the test are minimised**.

Principles of good practice in remotely delivered language assessment

Any high-quality language assessment and its associated practices should be grounded in a robust framework of principles. Central to this framework is the concept of **validity**, which can be defined as the extent to which the interpretations and decisions based on test results are well-founded and appropriate, supported by both empirical data and theoretical rationale (Messick, 1989). To make informed, holistic judgements about validity, it is essential to consider multiple dimensions and to gather various forms of evidence, both theoretical and empirical. Throughout this evaluative process, the principle of **fairness** must also be a key consideration, ensuring that the assessment is free from bias and equitable at every stage of the testing process (Walters, 2022; Xi, 2010).

Regardless of the test format or type, the decisions based on the test scores need to be valid for the intended purpose of the test. Beyond these essential dimensions contributing to **fitness for purpose**, the unique characteristics of remote assessment call for additional validity and fitness-for-purpose considerations. These considerations can be grouped into **six key areas, which are the basis for foundational principles** supporting the robustness of remotely delivered language assessments.

Figure 1. Six key principles for best practice in delivering remote language assessments



Six foundational principles and guiding questions

Construct coverage	What types of tasks are used to evaluate language proficiency in a remote setting, and are they fit for the purpose of the test?	Remote delivery should not compromise which aspects of language proficiency are assessed or how they are assessed. The test design and the tasks included must be carefully aligned with the specific language skills being evaluated and the overall purpose of the assessment.
Security	What measures are in place to ensure test security and to effectively detect instances of malpractice in remote assessment?	Establishing and maintaining trust and security in the remote testing process and, consequently, in the test results, is critical for remotely delivered assessments, particularly when used for high-stakes purposes.
Standardised administration conditions	To what extent is the assessment delivered consistently across different testing contexts and test-takers?	Each remotely delivered assessment session should follow consistent standards to ensure fairness for all test-takers. Since certain administrative aspects in remote settings fall outside the test provider's direct control, specific measures must be implemented to maintain consistency across sessions.
Degree of human involvement	What is the role of human input before, during, and after the remote assessment, and how does it support the testing process?	Appropriate human guidance, support, and oversight must remain integral to remotely delivered assessments, with AI used to assist, not replace, human judgement, particularly in high-stakes and legally regulated contexts.
Fair and inclusive participation	How accessible is the remote assessment to diverse groups of test-takers, and what steps are taken to promote equity in participation?	Remote delivery assessments need to provide not just geographic accessibility but also incorporate measures to accommodate unique test-taker circumstances whenever possible.
Comparability	How is comparability ensured across different modes of delivery when remote assessment is offered as an equivalent option to other modes of the same test?	When remote assessments are presented as an equivalent alternative to in-centre options, it is essential to validate and monitor the comparability of test quality and outcomes to ensure that scores from different delivery modes can be used interchangeably.

All of these areas of consideration and best practice need to be made publicly available, allowing stakeholders to review the evidence and make informed decisions about the use of remotely delivered assessments, as well as how to appropriately interpret and apply the resulting scores for their intended purposes.

Construct coverage

What types of tasks are used to evaluate language proficiency in a remote setting, and are they fit for the purpose of the test?	Remote delivery should not compromise which aspects of language proficiency are assessed or how they are assessed. The test design and the tasks included must be carefully aligned with the specific language skills being evaluated and the overall purpose of the assessment.

While test design is a fundamental consideration for any assessment, it becomes even more critical in remote settings. Practical benefits of remote assessment should not undermine what the test is designed to assess or its suitability for its intended purpose. **Relying solely on task types that are easy to implement in a remote test format but are limited in their coverage might not provide a solid foundation for assessment, especially for high-stakes decisions such as admissions or immigration.**

Consider the example of a face-to-face speaking assessment. In traditional formats, the examiner and the test-taker are physically present in the same space. In a remote context, this co-location is not possible, prompting adjustments to the format of the test. One option is to have a remotely delivered test which relies on machine-delivered prompts. This increases efficiency but limits the ability to assess interactive speaking skills. In certain contexts, this approach is justifiable as it aligns with the purpose of the test. In other contexts, for example, for admission into higher education institutions, which require the demonstration of a breadth of language skills, such an approach might have limitations.

An alternative is a remote test which utilises online video-call technology. Video-conferencing technology has lent itself naturally to the delivery of real-time, interactive speaking assessments. For instance, the IELTS test administers the Video-call Speaking (VCS) test, an equivalent alternative to the in-person Speaking test, in which the test-takers and examiner join remotely. In this way, remotely delivered tests have helped **extend the availability of face-to-face speaking assessment** to geographically remote or politically sensitive areas where examiners may not be readily available. However, this still requires careful consideration of whether test-takers in these contexts can participate under equitable and appropriate conditions. It is important to note, though, that access to the necessary technology and internet may still be a barrier for some (Nakatsuhara et al., 2017; 2021).

Whichever approach is chosen, the **priority must be to ensure that the assessment remains aligned with its intended purpose. Practicality needs to be balanced against the need for a valid and meaningful evaluation of language proficiency,** and the appropriateness of the chosen test for its purpose.

Security

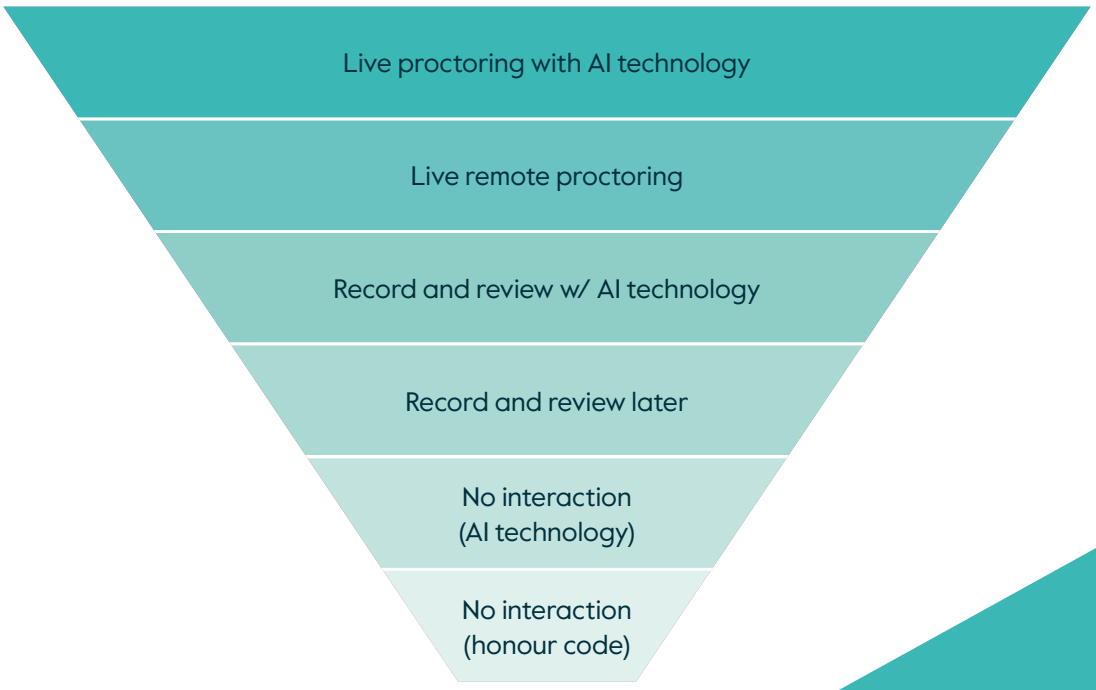
What measures are in place to ensure test security and to effectively detect instances of malpractice in remote assessment?	Establishing and maintaining trust and security in the remote testing process and, consequently, in the test results, is critical for remotely delivered assessments, particularly when used for high-stakes purposes.

Security is a major consideration for all high-stakes tests, since they often involve life-changing decisions for test-takers. Remote assessment needs to place even greater emphasis on maintaining rigorous test security, since it is more vulnerable to exam malpractice.

Due to the physical separation between test-takers and invigilators, remote proctoring becomes necessary in remote assessment contexts in order to maintain a high standard of security. **Remote proctoring** – i.e., the process for overseeing the integrity of the test and supporting test-takers – is critical to ensure the integrity of the testing process, alongside robust technical controls such as a secure lockdown browser and strong ID verification.

Remote proctoring can be implemented with different **levels of security**, depending on the purpose and stakes of the test, as seen in the hierarchy of remote proctoring options.

Figure 2. Hierarchy of remote proctoring options (adapted from Michel, 2020)



While live remote proctoring enables real-time intervention and is therefore often selected for high-stakes assessments, record-and-review proctoring also offers a robust and scalable approach to maintaining test integrity. In this option, sessions are recorded and subsequently reviewed by proctors, often supported by AI-generated alerts, and the approach offers notable advantages in terms of practicality, cost-effectiveness, and operational scale. Importantly, the most appropriate configuration depends on the risk profile of the test and the security demands associated with its intended use.

The proctoring approaches are supported by a range of technical tools and monitoring methods, which can be combined in different ways depending on the required level of security. These tools can range from simple lockdown browsers that prevent leaving the test window, thereby minimising some forms of malpractice, to more advanced artificial intelligence (AI) applications, such as AI-monitored logs of user computer input, including key presses and mouse movements, and computer vision technology used to identify suspicious behaviour via webcam. In many systems, additional measures such as a second camera angle (e.g., via a mobile device) could provide a more comprehensive view of the testing environment. With different combinations of these security measures, the timing of monitoring can also vary. In some cases, the entire test session is recorded and reviewed after the test (record and review later or with AI technology); in others, live monitoring allows for real-time intervention if suspicious activity is detected (live remote proctoring; live proctoring with AI technology). Post-test analyses, including statistical irregularity detection, biometric verification, and scrutiny of score patterns, remain a crucial part of a layered security approach, as no single proctoring configuration can prevent or detect all forms of misconduct.

Currently, different test providers apply various combinations of these measures, leading to a range of possible scenarios for remote proctoring (Isbell & Kremmel, 2020). Various proctoring tools, ranging from technology-enabled to human-managed, and different proctoring timings, such as recorded versus live, are available in remote delivery, allowing for a range of implementation scenarios. However, not every combination guarantees the same level of effectiveness in ensuring security or reliably detecting malpractice. For instance, live remote proctoring with concurrent session recording is often considered superior to record-and-review (asynchronous) proctoring alone, as it allows real-time intervention and also provides a full recording for later review, thereby reducing the risk of exam content being harvested or exposed. In some implementations, however, human intervention is not continuous: automated monitoring may flag sessions and prompt targeted human review or live escalation only when specific risks are detected.

In all cases, the level of remote proctoring needs to be fit for the test purpose and context. In high-stakes cases, e.g., for certification or admissions purposes, the most secure level of proctoring is needed, and in many cases, in-centre delivery may be more appropriate than remote.

As remote delivery evolves, the nature of security risks has also shifted. Certain forms of exam malpractice that were more common in paper-based or fixed-form testing, such as copying the responses of someone nearby, have become far less feasible with the widespread use of computer-delivered tests, randomised item presentation, and adaptive testing. However, remote delivery introduces new and evolving challenges, such as illegally recording or copying test content (also known as 'item harvesting'), using unauthorised resources during the test, or various forms of impersonation. These may involve the use of hard-to-authenticate IDs, identity spoofing techniques, or hidden external assistance (e.g., an imposter inside or outside the room providing answers via wireless devices). These risks place increased demands on effective monitoring and post-test scrutiny to ensure that results are valid, secure, and trustworthy.

Recent advances in generative AI have introduced a host of additional risks. For example, they have made it increasingly feasible to create highly convincing deepfake audio or video, which could undermine the reliability of remote identity verification processes if appropriate safeguards are not in place. In addition, test-takers may attempt to use covert connected devices, such as hidden earpieces, Bluetooth or wired communication tools, or secondary devices positioned outside the webcam's field of view, to obtain unauthorised assistance during the test. These evolving forms of malpractice highlight the need for continuous monitoring, regular updates to security protocols, and ongoing research to ensure that remote assessments remain robust against both current and emerging threats.

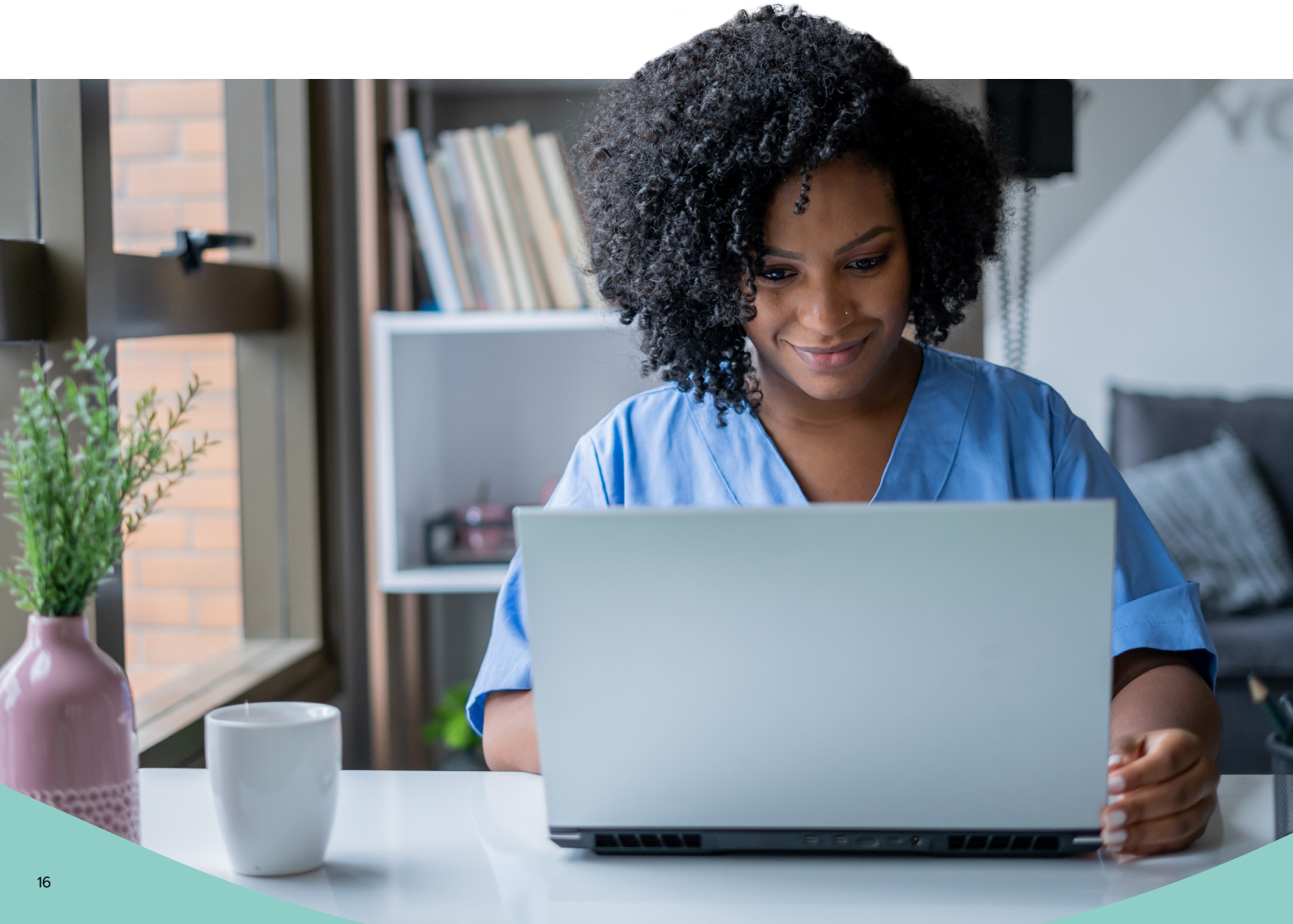
In terms of unique benefits, remote delivery facilitates the collection and analysis of detailed test-taking records, which can help monitor test security practices, identify potential misconduct, and support test-takers in appealing administrative decisions.

Standardised administration conditions

To what extent is the assessment delivered consistently across different testing contexts and test-takers?	Each remotely delivered assessment session should follow consistent standards to ensure fairness for all test-takers. Since certain administrative aspects in remote settings fall outside the test provider's direct control, specific measures must be implemented to maintain consistency across sessions.

Test-takers should be clearly informed in advance about how to set up their chosen testing location to meet required standards and maintain those conditions throughout the test. Their active cooperation, in tandem with the test provider's efforts, is essential for achieving a standardised testing environment and ensuring fairness for all.

From the test provider's perspective, the **testing platform should be technically light** where possible, requiring minimal computing power and technical expertise from users. However, some security configurations, such as dual camera setups or continuous AI-assisted monitoring, inevitably demand higher bandwidth. Clear guidance and pre-test checks are therefore essential to ensure that test-takers' equipment and internet connections can support the required proctoring measures. A user-friendly and streamlined platform helps minimise disruptions caused by system or procedural complexities and supports the delivery of consistent assessment sessions and equitable testing conditions.



Degree of human involvement

What is the role of human input before, during, and after the remote assessment, and how does it support the testing process?	Appropriate human guidance, support, and oversight must remain integral to remotely delivered assessments, with AI used to assist, not replace, human judgement, particularly in high-stakes and legally regulated contexts.

While advancements in technology and AI have made it increasingly feasible to administer remote assessments without human involvement, **human input is indispensable,** particularly in high-stakes contexts where human oversight must remain central. From setting up the appropriate test environment to overseeing the testing process, human involvement streamlines procedures and helps ensure secure, fair, and effective outcomes at every stage.

Human assistance, whether in the form of a support channel for addressing pre-test requirements or a live proctor during the test, can help create a testing environment that supports security and enables conditions that allow test-takers to perform at their best.

Incorporating a layer of live monitoring by human proctors into the remote proctoring system can further strengthen malpractice detection and, in turn, enhance the credibility of the test results. Moreover, live proctors can offer real-time support for technical or procedural issues, helping to ensure a consistent and uninterrupted assessment experience. For instance, human involvement can help address linguistic challenges related to test procedures, such as general instructions and security protocols, which may be particularly difficult for lower-proficiency test-takers in remote settings.

To maximise the benefits of human involvement, human proctors must be carefully vetted, well-trained, and supported through ongoing quality-assurance processes to ensure they can deliver assistance effectively and take appropriate actions consistently. While this may inevitably increase costs, striking a balance between test integrity and practicality is essential. It is also essential to ensure that proctors act professionally, given their access to test-takers and sensitive data. One advantage of remote delivery is that proctor performance can often be more easily monitored and reviewed, thereby enhancing quality assurance.

Fair and inclusive participation

How accessible is the remote assessment to diverse groups of test-takers, and what steps are taken to promote equity in participation?	Remote delivery assessments need to provide not just geographic accessibility but also incorporate measures to accommodate unique test-taker circumstances whenever possible.

Remotely delivered assessments can reach a larger group of test-takers, including those who may have been harder to reach through traditional test administrations or ones with accessibility challenges, such as those living in remote areas or those with mobility difficulties, offering more flexibility in test scheduling.

Remote assessment can also assist in better accommodating test-takers with certain disabilities. For example, taking the test in a familiar environment can benefit neurodivergent individuals or those with anxiety by helping to reduce overstimulation and stress. However, some features of remote testing may inadvertently disadvantage specific groups. For instance, those with involuntary eye movements may be unfairly flagged if eye gaze direction is monitored as part of the remote proctoring process to detect potential cheating or distractions. Careful planning is required to ensure that no group is unfairly disadvantaged. As remote test delivery is still a relatively new operational mode, the full range of appropriate access arrangements has not yet been comprehensively explored, validated, or implemented. For this reason, it may be **premature to rely solely on remote delivery without a robust understanding of how accommodation needs can be met equitably across all test-taker groups.**

The technological divide is another barrier to inclusive access, as remotely delivered assessments require test-takers to have their own devices and a stable internet connection. Relying solely on remote assessments as the only testing option may not bridge this divide in the foreseeable future, so offering alternative modes of test delivery, e.g., in-centre and online, which are comparable, may better accommodate different individual circumstances.

Economic accessibility should be considered, too. Remote assessments can reduce travel costs for many test-takers. At a minimum, remotely delivered assessments should not be more expensive than those offered at test centres.

Comparability

How is comparability ensured across different modes of delivery when remote assessment is offered as an equivalent option to other modes of the same test?

When remote assessments are presented as an equivalent alternative to in-centre options, it is essential to validate and monitor the comparability of test quality and outcomes to ensure that scores from different delivery modes can be used interchangeably.

Currently, many test providers offer remote assessments as a comparable alternative to in-centre testing. To support this comparability, it is essential to validate that the remote option meets the same level of assessment standards and that the scores it produces are comparable to those from in-centre assessments. This supports the equivalent interpretation of test results across delivery modes, enabling appropriate use for their intended purposes. Such validation should focus on a range of aspects across delivery modes, including the comparability of scores, the cognitive processes and language elicited, and the testing conditions.

It is crucial that this validation is not treated as a one-time effort. Continuous monitoring is necessary to assess the ongoing comparability of different delivery modes.

For further information on remotely delivered assessment and associated concerns, please refer to the following publications:

- 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
- Isbell, D. R., Kremmel, B., & Kim, J. (2023). Remote proctoring in language testing: Implications for fairness and justice. *Language Assessment Quarterly, 20*(4-5), 469–487. https://doi.org/10.1080/15434303.2023.2288251
- Moses, T., & Puhan, G. (Eds.). (2022). Maintaining score comparability [Special issue]. *Journal of Educational Measurement*, 59(2). https://doi.org/10.1111/jedm.12319
- Muhammad, A. A., & Ockey, G. J. (2021). Upholding language assessment quality during the COVID-19 pandemic: Some final thoughts and questions. *Language Assessment Quarterly, 18*(1), 51–55. https://doi.org/10.1080/15434303.2020.1867555
- Nakatsuhara, F., Inoue, C., Berry, V., & Galaczi, E. (2017). Exploring the use of video-conferencing technology in the assessment of spoken language: A mixed-methods study. *Language Assessment Quarterly, 14*(1), 1–18. https://doi.org/10.1080/15434303.2016.1263637

From principles to practice: A Cambridge approach to remote language assessment

In this section, we illustrate how the principles of good practice are translated into practice in the remotely delivered mode of Cambridge English tests, drawing on specific examples ranging from proof-of-concept initiatives to tests currently in operation.

The Cambridge portfolio includes both remote-first tests such as Linguaskill and remote versions of established in-centre assessments such as IELTS Online. Tests such as these will serve as representative examples, demonstrating how these principles have been integrated into the planning, administration, and validation processes for remote assessment.

Construct coverage – How does Cambridge maintain the integrity of what is being assessed in remote language assessments?

At the heart of all Cambridge tests is communicative language testing, which aims to evaluate not just what test-takers know about the language, but how effectively they can use it to achieve real-world communication goals. Migrating the existing tests to remote delivery should not compromise this core foundation of the assessment, even though such a transition may involve challenges that must be carefully addressed.

We will take the Speaking component of the Cambridge English Qualifications (CEQs) as an example. One of its distinctive features is the inclusion of a variety of tasks that elicit different types of speech, including both spoken production and interaction, and engage a range of cognitive processes accordingly. This is often facilitated by a human examiner who conducts the test using a structured interlocutor frame. Preserving this format in a remote context requires a way to connect multiple participants in different physical locations into a shared virtual space where interactions similar to those that occur in person can also take place online. A similar approach can also be seen in the Speaking component of IELTS delivered via video calls (Nakatsuhara et al., 2017; 2021).

The possibility of migrating the Speaking component of the Cambridge English Qualifications (CEQs) to a remote context has been explored as a proof-of-concept, involving a pair of test-takers and two examiners. This was successfully piloted with the A2 Key, B1 Preliminary, and B2 First exams (Galaczi & Lee, 2023; Lee et al., 2024). (At the time of writing, CEQ Speaking remains in person; this programme of work is intended to inform any future operational roll-out of video-call delivery.)

The remotely delivered CEQ Speaking tests retain the same content and structure as the in-person mode. However, slight modifications to the delivery are inevitable to cater to the unique features of the remote format. For instance, non-verbal cues in a video call are interpreted slightly differently, as they are transmitted through a screen. This can impact how an examiner might use hand gestures to prompt a test-taker to look at a task. Additionally, some parts of the test involve visual prompts designed to elicit specific types of utterances. Since these materials cannot be directly shown to the test-takers, they are instead shared via screen, which necessitates careful consideration of the visibility of the content when shared. These changes require careful planning for the test provider and additional training and practice for the examiners to ensure they are familiar with, and confident in, the adjustments.

In a contrasting example, Linguaskill was designed as an online English language test without a face-to-face, examiner-led mode, to be taken remotely or in a test centre. Unlike some remotely delivered assessments where the online format requires adjustments to administration procedures, Linguaskill uses exactly the same test regardless of whether it is taken remotely or in a test centre. The same adaptive engine, task types, scoring processes, and test construct apply across delivery modes. Only the delivery infrastructure and security model differ. This ensures that candidates receive an equivalent test-taking experience and that scores carry the same interpretation across modes.

Returning to Linguaskill, the Speaking component operates differently from examiner-led speaking tests. Instead of a live interlocutor, pre-recorded audio prompts guide candidates through a series of tasks, each carefully designed to elicit different elements of speaking, such as discussing, explaining, and justifying. The Linguaskill Speaking component, which covers CEFR B1–C2 levels, aims to assess real-world communicative skills through a variety of visual and audio inputs. Candidates first listen to, read, or view visual materials (as illustrated in the figure below), then select relevant information to complete tasks requiring various communication aspects such as summarising, recommending, or speculating skills. Within the test, candidates encounter a mix of input materials; for example, in Part 3, candidates are given an infographic (see the figure below) and are asked to provide a recommendation to a friend. Delivered online, this format enhances flexibility and supports the use of varied multimodal inputs to elicit a range of communicative functions, while targeting different aspects of speaking from paired, interaction-based formats.

Figure 3. Example of visual material presented in a Linguaskill Speaking test

Security – How does Cambridge ensure a high level of test security?

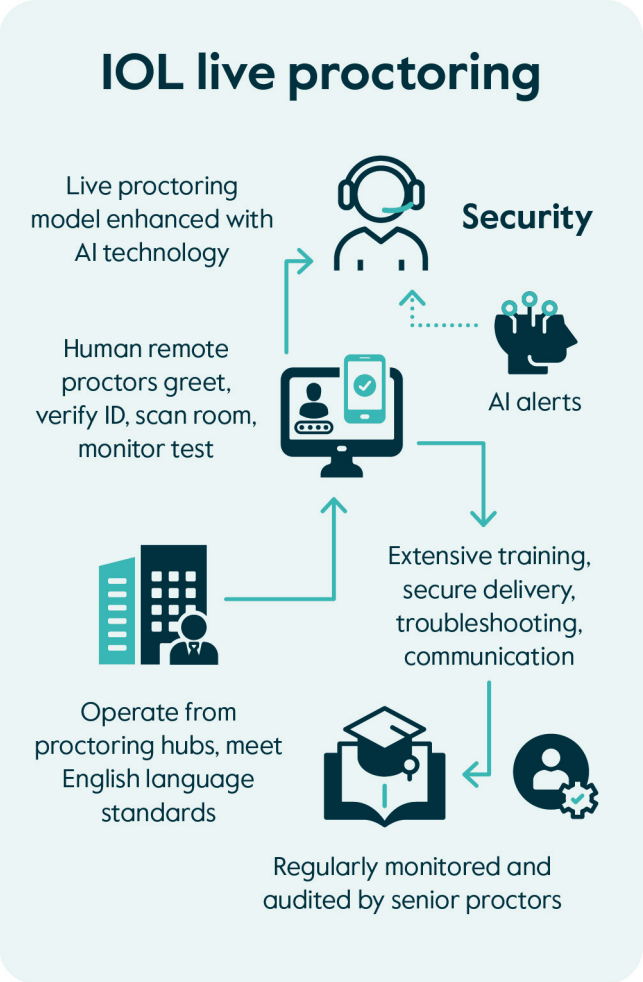
As highlighted throughout this paper, when scores from remotely delivered assessments are used, particularly for high-stakes purposes, it is essential to ensure a high level of test security. This is necessary to confirm that the scores accurately reflect the language proficiency of the individual registered to take the test, without any breach of security protocols. An important component of this is a robust remote proctoring system, complemented by technical safeguards (e.g., a secure lockdown browser, robust ID verification, and controls to mitigate risks such as remote access software and unauthorised devices), designed to uphold a level of rigour comparable to that found in test centres.

An example of a Cambridge English test with stringent security measures is IELTS Online (IOL) – a remotely delivered version of IELTS Academic launched in 2022. Its rollout was gradual, beginning in a few select locations and expanding globally in order to minimise risk and ensure that the technological capabilities are robust enough to support large-scale delivery of this high-stakes assessment.

IOL employs a live proctoring model enhanced with AI technology, which sits at the top of Michel's (2020) security hierarchy for remote proctoring. This hybrid approach pairs human oversight with AI-driven flagging, providing an added layer of security. During check-in, a trained Greeter welcomes the test-taker, verifies identity, and completes an environment check, including a 360-degree room scan, before the test-taker is released into the test session. The live proctor then monitors the test session in real time, supported by AI software that flags potentially suspicious behaviour for review and follow-up where necessary.

IOL proctors operate from dedicated proctoring hubs and are required to meet specific English language proficiency standards. They undergo extensive training covering secure test delivery, technical troubleshooting, effective communication with test-takers, and more. Before overseeing live test sessions, proctors participate in mock test scenarios featuring potential security breaches, and shadow live test events for hands-on experience. Their performance in live proctoring is also regularly monitored and audited by experienced senior proctors.

Figure 4. IOL live proctoring process and security measures



The combination of live human proctoring and technology-driven alerts effectively identifies and addresses malpractice, with each contributing distinct advantages. Technology excels at detecting subtle or otherwise invisible forms of cheating, such as unauthorised software use, and consistently flags behaviours like gaze aversion or background noise. However, it cannot interpret the nuanced context of human behaviour. While it can indicate potentially suspicious activity, human judgement is essential to determine whether such behaviours genuinely constitute malpractice. Human proctors also provide the flexibility to respond to unexpected situations beyond the limits of programmed systems. Together, they form a robust system for safeguarding test integrity in IOL.

Standardised administration and human involvement – How does Cambridge support remote test-takers in creating a testing environment that meets required standards?

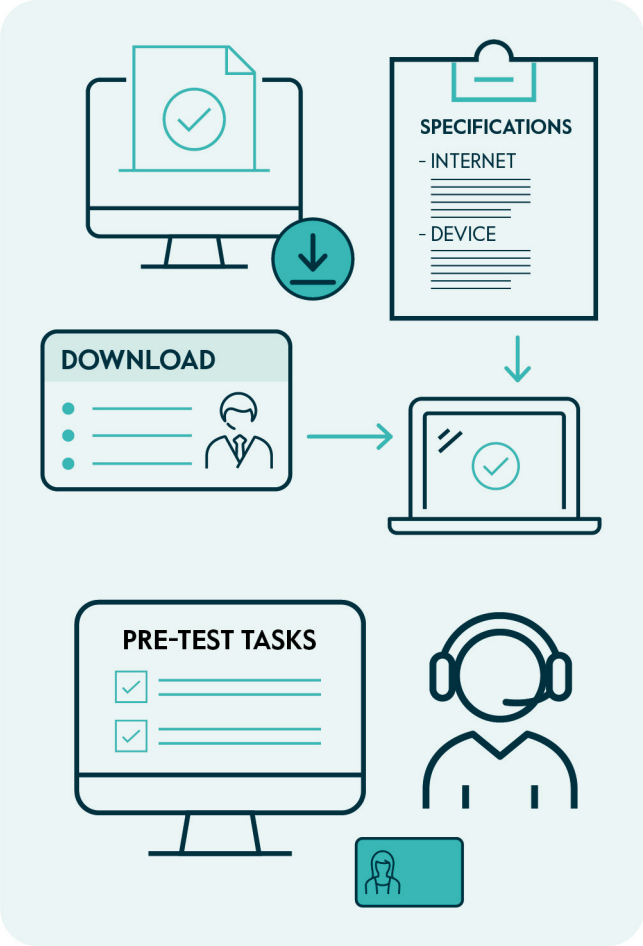
While robust remote proctoring plays a critical role in maintaining test integrity, the active cooperation of remote test-takers in setting up an appropriate, standardised test environment is equally important. This can be supported through clear, accessible guidance provided before the test and responsive assistance during the test session. Providing guidance in multiple formats and at various stages of the test journey fosters a collaborative effort with test-takers in creating a compliant testing environment and upholding test integrity. This principle is one of the central pillars of the Cambridge approach to remote assessments and is actively implemented in current test offerings, with plans for continued application in future ones.

Using IOL as an example, registered test-takers receive an email containing essential information about the testing process shortly after their registration. This includes access to preparation materials, familiarisation tests, and a detailed overview of the technical requirements, such as the minimum and recommended specifications for operating systems, internet connections, and devices.

To ensure their setup meets these requirements, test-takers must download a dedicated software tool onto the computer they plan to use. This tool provides step-by-step instructions and runs system diagnostics to confirm compliance. The email also outlines a timeline of key pre-test tasks, including conducting a room scan and checking valid identification for verification, that test-takers should be aware of in advance.

Information is delivered in multiple formats such as text, video, and across various communication channels to enhance accessibility and engagement. The test-taker portal also offers support features like live chat and a request form for additional help with individual concerns not covered by the resources.

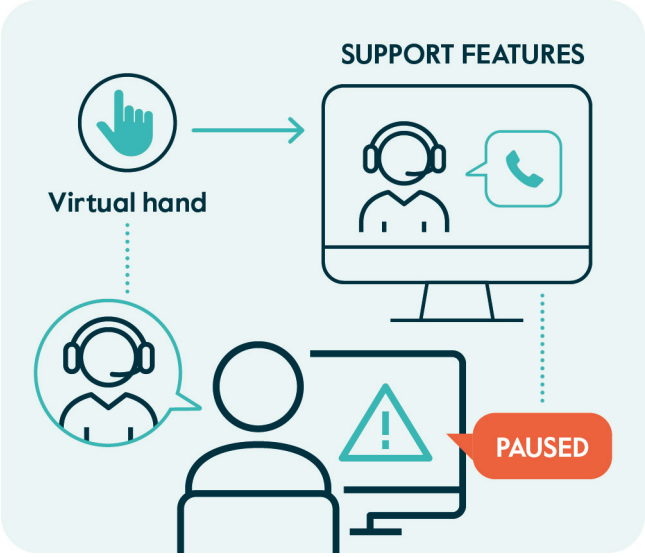
Figure 5. Technical and human support touchpoints in IOL



Responsive support is also available during the live test session via a human proctor. While the room scan occurs before the test begins, proctors can provide real-time guidance if test-takers have missed any requirements outlined in the pre-test instructions. They help resolve any missing elements or adjust improper setups to ensure a secure environment before the test starts.

If assistance is needed during the test, test-takers can raise a virtual hand or use the chat function within the testing platform. Proctors can pause the test and offer appropriate administrative support through live video. Beyond their monitoring duties, proctors also recognise and embrace their role in supporting test-takers in co-creating a secure and comfortable testing environment (Bruce & Clark, in press).

Figure 6. Real-time support features in IOL



Inclusive access – How does Cambridge promote broader access through remote language assessment?

A key rationale for expanding into remote assessment is to make assessment available to test-takers who are unable to attend a testing centre due to individual circumstances. However, broader availability does not in itself guarantee equitable access, which depends on whether remote conditions allow fair and meaningful participation. In addition to offering convenience, such as more flexible scheduling, remote assessment can reduce the need to travel to a test centre that may be hundreds of miles away and may also lower indirect costs. Together, these factors can create opportunities for individuals who might otherwise be unable to take the test at all.

Appropriate special arrangements are available to **accommodate those with specific needs**. The precise features vary by delivery platform. For example, IELTS Online offers options such as enlarged text and alternative colour contrast settings. Other platforms, such as those used for Linguaskill, support accessibility in different ways, for instance, through pause and play audio controls in listening tests. Systematic research is underway to determine which further specific needs can be reliably supported across platforms. Until such evidence is available, remote delivery may not be suitable for all high-stakes contexts or for test-takers requiring extensive accommodations. In these cases, in-centre delivery continues to be the preferred and recommended mode, ensuring full access to the range of resources, security procedures, and controlled conditions required.

Another consideration regarding inclusive access is whether test-takers have the necessary technology and home settings to create suitable remote testing environments. Not every test-taker has access to the full range of resources or the capability to use them to meet the suggested requirements. To address this, remotely delivered Cambridge assessments are offered as just one of the options a test-taker can choose, alongside other delivery modes such as paper-based or computer-based tests administered at test centres. The availability of alternative delivery modes ensures that no group of test-takers is excluded; offering flexibility helps address potential gaps in technology access and personal preferences.

Comparability – How does Cambridge determine whether the remote mode of test delivery is fit for purpose and aligns with the existing testing programme?

As important as it is to put principles of good practice into action, it is equally essential to evaluate whether implementation occurs as intended. Cambridge supports this through continuous research conducted both before and after the operationalisation of remotely delivered tests and, where appropriate, disseminates findings through publicly available reports and research publications to promote transparency (e.g., Bruce & Clark, in press; Galaczi & Lee, 2023; Lee et al., 2024; Lee & Tasviri, in press; Nakatsuhara et al., 2017, 2021).

A core component of this work is validating the comparability of remote assessment formats with other delivery modes, so that the constructs assessed, and interpretations of resulting scores, remain consistent across modes. For example, research comparing the remotely delivered video-call Speaking test with the in-person Speaking test administered in test centres has examined multiple sources of evidence, including scores, language functions elicited, rating processes, and the perceptions of test-takers and examiners (Galaczi & Lee, 2023).

Beyond the test design phase, ongoing operational monitoring is also vital. One key element involves gathering feedback from critical stakeholder groups, such as test-takers, examiners, score users, and remote proctors, on how the test is administered. Their perceptions and experiences have been examined systematically, with particular attention to impacts on test performance and administration, the effectiveness of proctoring support, and the robustness of security measures. Evidence from these studies has informed targeted refinements (e.g., to test-taker guidance and proctor training), illustrating how feedback helps maintain fitness for purpose in remote delivery (Bruce & Clark, in press; Bruce & Lee, 2023; Lee & Tasviri, in press). Collectively, these research and monitoring activities strengthen the validity argument for remotely delivered tests and support evaluation of their comparability with established test delivery modes.

The future of remotely delivered language assessment

Remotely delivered language assessment is now a viable alternative alongside traditional test centre-based options. While the key advantages of remote delivery – **convenience and flexibility** – are undeniable, these **benefits must not compromise fitness for purpose, particularly when test scores are used to inform high-stakes decisions such as those related to higher education admission or immigration**. To maintain the integrity of such assessments, the principles outlined in this paper should be applied at every stage of testing practices.

The success of remotely delivered language assessments relies on the collaborative efforts of all parties involved. Creating optimal and fair testing conditions is not the sole responsibility of test providers; it also requires active participation from test-takers (Lee & Tasviri, in press). This shift reinforces the importance of implementing the use of remote vs in-centre testing based on robust considerations of its appropriateness to specific situations, and the fact that in some high-stakes situations, a remote approach may not be the most optimal approach. It is essential that fitness for purpose is assessed for each context. Remote delivery may not be appropriate for certain scenarios, and in-centre delivery might offer a better fit with the purpose of the assessment.

Looking ahead, the rapid advancement of AI technology is poised to further reshape the landscape of remotely delivered language assessment, with several aspects of this already being explored. In terms of test design, a conversational AI agent could be employed to deliver remote speaking tests, enhancing convenience while maintaining the interactive nature of the assessment. This approach would allow interactive speaking tasks to be included even in a fully automated test design, provided that the interaction can be shown to be sufficiently natural (Karatay & Xu, 2025). In remote proctoring, the ongoing improvements and integration of AI applications are expected to increase the consistency and accuracy of malpractice detection, thereby enhancing test security.

Furthermore, as advanced technology becomes more widely integrated into individuals' daily and professional lives and more affordable worldwide, test-takers are likely to become increasingly familiar with digital contexts, including remote assessment. This growing familiarity can enhance their confidence in choosing to take tests remotely and meeting the requirements.

All of these developments solidify the advantages of remote test delivery while addressing current or potential challenges. Central to this is the goal of ensuring that remote tests are **fit for purpose**, supported by an extensive array of research evidence. Even with a high level of confidence in a robust system for safeguarding test integrity, a cautious approach remains essential. There may be security vulnerabilities that are not yet fully understood or observable in real time. For this reason, the use of scores from remotely delivered assessments for high-stakes decisions should be considered with particular care and should not prematurely replace established in-centre delivery modes. Ultimately, where an evaluation of the advantages and limitations of remote delivery indicates that its risks cannot be sufficiently mitigated for a given high-stakes context, in-centre testing should remain the preferred mode of delivery. Remote delivery should only be adopted once its security and operational challenges can be addressed to a level that ensures it is fit for the intended purpose.

References

Association of Test Publishers, & National College Testing Association. (2024, May 1). *Assessment industry standards and best practices for the online observation of tests* (Public comment draft). Available online.

Bruce, E., & Clark, T. (in press). Remote proctoring and implications for the test-taker experience: Using a theory of change model for evidence-based global rollout. In T. Clark & Bruce, E. (Eds.), *Quality standards for remote digital language testing* (pp. 12–23). Palgrave Macmillan.

Bruce, E., & Lee, H. (2023). *Preliminary test-taker perspectives of a high-stakes assessment conducted online* [Conference presentation]. International Association of Teachers of English as a Foreign Language, Harrogate, United Kingdom.

Burgess, B., Ginsberg, A., Felten, E., & Cohney, S. (2022). Watching the watchers: bias and vulnerability in remote proctoring software. In *Proceedings of the 31st USENIX Security Symposium* (pp. 571–588). https://arxiv.org/abs/2205.03009

Galaczi, E., & Lee, H. (eds.). (2023). Exploring video-call paired speaking tests [Special issue]. *Research Notes, 86*. Available online.

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

Karatay, Y., & Xu, J. (2025). Exploring the potential of conversational AI for assessing second language oral proficiency. *TESOL Quarterly, 59*, S220–S250. https://doi.org/10.1002/tesq.70003

Lee, H., & Tasviri, R. (in press). Validating an at-home English proficiency test: Voices of test-takers and examiners. In T. Clark & Bruce, E. (Eds.), *Quality standards for remote digital language testing* (pp. 37–49). Palgrave Macmillan.

Lee, H., Mullooly, A., Devine, A., & Galaczi, E. (2024). Exploring interaction in video-call paired speaking tests: A look at scores, language, and perceptions. *Applied Linguistics, 45*(3), 533–552. https://doi.org/10.1093/applin/amad046

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan.

Michel, R. S. (2020). Remotely proctored K-12 high stakes standardized testing during COVID-19: Will it last? *Educational Measurement Issues & Practice, 39*(3), 28–30. https://doi.org/10.1111/emip.12364

Nakatsuhara, F., Inoue, C., Berry, V., & Galaczi, E. (2017). Exploring the use of video-conferencing technology in the assessment of spoken language: A mixed-methods study. *Language Assessment Quarterly, 14*(1), 1–18. https://doi.org/10.1080/15434303.2016.1263637

Nakatsuhara, F., Inoue, C., Berry, V., & Galaczi, E. (2021). Video-conferencing speaking tests: Do they measure the same construct as face-to-face tests? *Assessment in Education: Principles, Policy & Practice, 28*(4), 369–388. https://dx.doi.org/10.1080/0969594X.2021.1951163

Ofqual. (2024, April 24). *Ofqual's approach to regulating the use of artificial intelligence in the qualifications sector.* Available online.

Terpstra, A., De Rooji, A., & Schouten, A. (2023). Online proctoring: Privacy invasion or study alleviation? Discovering acceptability using contextual integrity. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–20). https://doi.org/10.1145/3544548.3581181

Walters, F. S. (2022). Ethics and fairness. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (pp. 563–577). Routledge.

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing, 27*(2), 147–170. https://doi.org/10.1177/0265532209349465

Yoder-Himes, D. R., Asif, A., Kinney, K., Brandt, T. J., Cecil, R. E., Himes, P. R., Cashon, C., Hopp, R. M. P., & Ross, E. (2022). Racial, skin tone, and sex disparities in automated proctoring software. *Frontiers in Education, 7*, 881449. https://doi.org/10.3389/feduc.2022.881449

Author biographies

Hye-won Lee is a Senior Research Manager at Cambridge University Press & Assessment, where she leads research projects on next-generation language assessments and provides research-based evidence to validate the quality of current and future assessments. She holds a PhD in Applied Linguistics and Technology from Iowa State University, USA. Her research interests include remotely delivered assessment, language assessment literacy for recognising organisations, and the role of generative AI in learning-oriented assessment.

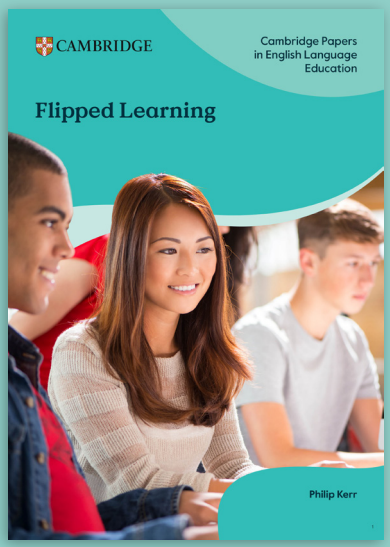
Daniel R. Isbell is an Associate Professor in the Department of Second Language Studies at the University of Hawai`i at Mānoa where he teaches courses in language assessment and quantitative research methods. He holds a MA in Teaching English as a Second Language (Northern Arizona University, USA) and a PhD in Second Language Studies (Michigan State University, USA). His research interests include second language speaking assessment, diagnostic language assessment, and the use of language assessments in higher education.

Benjamin Kremmel is a senior scientist at the University of Innsbruck, Austria, and the head of Language Testing Research Group Innsbruck (LTRGI). He holds an MA in Language Testing from Lancaster University, UK and a PhD in Applied Linguistics from the University of Nottingham, UK. He was the guest co-editor (with Daniel Isbell) of the 2024 Language Testing Special Issue on "Open Science Practices in Language Assessment," and is the series co-editor (with Paula Winke) of Applied Linguistics Press' Syllabi in Applied Linguistics. He is an EALTA Expert Member and the recipient of the 2023 TOEFL Essentials New Scholar Award.

Explore more

Cambridge Papers

in English Language Education



cambridge.org/english-insights

Find out more at
cambridge.org

We believe that English can unlock a lifetime of experiences and, together with teachers and our partners, we help people to learn and confidently prove their skills to the world.

Where your world grows

CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

All details are correct at the time of going to print in April 2026.

AHDRES/42224/0326