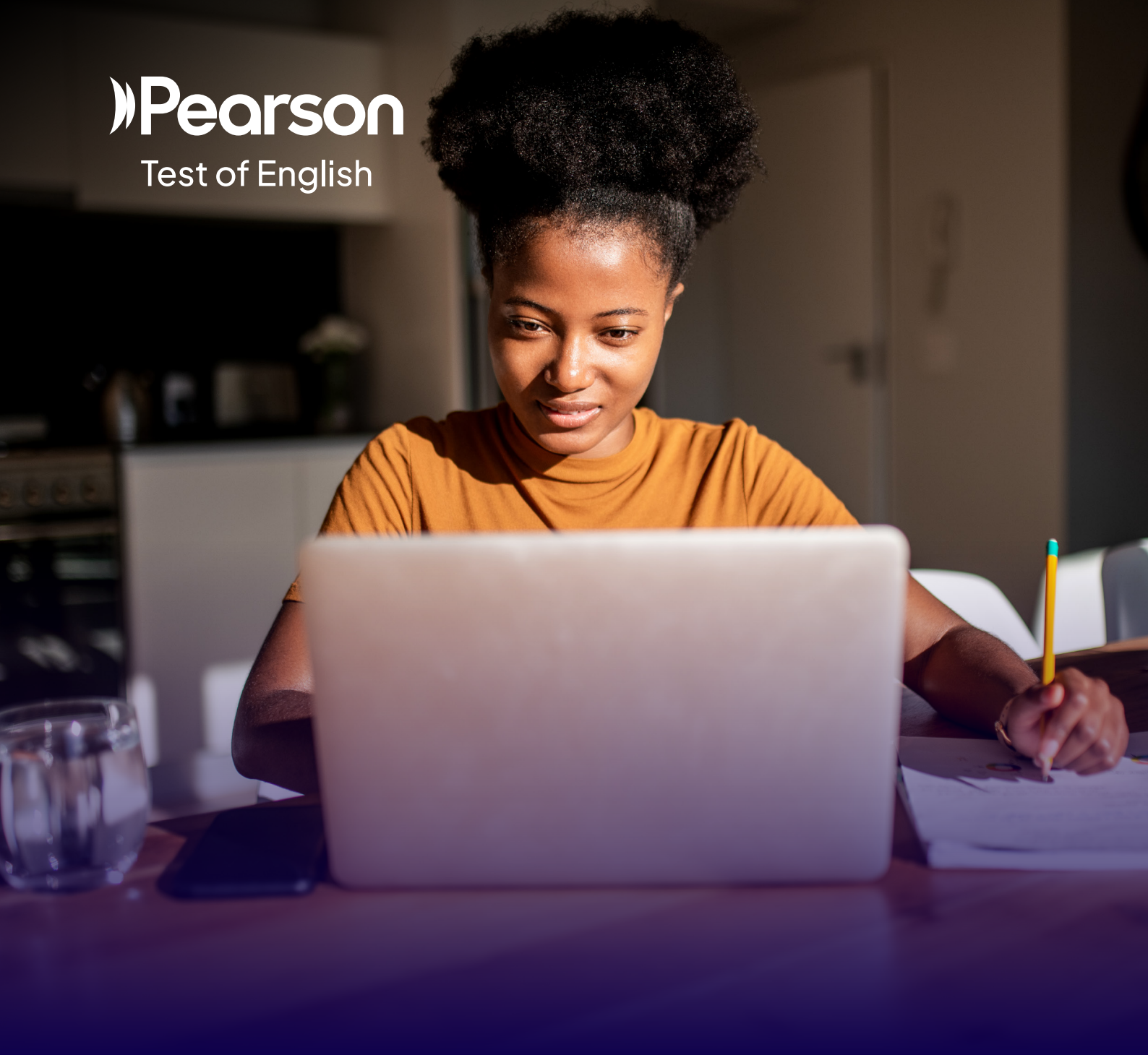




Ellen Barrow

Test Taker Perspectives

PTE Academic Enhancements

Contents

Background	3
Introduction	4
Methodology	5
Findings	6
Test Taker Background	6
Test and Question Level Feedback	9
Feedback on New Question Types	11
Test Preparation	17
Attitudes to AI Scoring in High-Stakes Tests	19
Conclusion	23

Background

Pearson Test of English (PTE) Academic is a computer-based international test of English.

PTE Academic was launched in 2009, in response to demand from higher education, governments and professional bodies for a test to measure the English communication skills of international students and economic migrants securely and accurately. The purpose of the test is to measure test takers' English language competency in listening, reading, speaking, and writing for academic and skilled migration purposes.

The foundation of PTE Academic was and continues to be the Common European Framework of Reference for Languages (CEFR, Council of Europe, 2001).

Research, expansion and elaboration of the CEFR has continued from the original 2001 publication, and it has undergone several key updates since its original publication, with the latest Companion volume published in 2020 (Council of Europe, 2020). Key changes to the CEFR promote more integration of mediation and interaction competencies into language learning and assessment.

It has therefore been an ongoing focus of PTE research to develop new question types and scoring rubrics that complement and expand the linguistic competencies of the test in line with CEFR enhancements.

“

It has been an ongoing focus of PTE research to develop new question types and scoring rubrics that complement and expand the linguistic competencies of the test in line with CEFR enhancements.

Introduction

As the landscape of language testing purposes continues to evolve, it is crucial to understand the perspectives of those who engage with these assessments.

Understanding test taker perspectives provides unique insights which enhance validity measures, contribute to test development, and increase trust and confidence in our assessments.

This research paper delves into the experiences and opinions of test takers who took part in a concordance study, taking the enhanced PTE Academic test and the IELTS Academic test within a 90-day period in early 2024. It provides a comprehensive analysis of their feedback on the revised PTE Academic test.

This research contributes to a wider program of work centered on the enhanced PTE Academic test, creating a broad range of evidence that validates and supports the updates made to the test. The most noticeable enhancements to test takers were the inclusion of two new speaking question types requiring extended spoken responses, purposed to reflect real-world language use: Respond to a Situation and Summarize Group Discussion. In this research phase, we sought feedback from test takers to evaluate the appropriateness and face validity of these items.

The primary aim of this study was to gather insights into the test takers' backgrounds, motivations, and their experiences of the two new question types, Respond to a Situation and Summarize Group Discussion. By examining these perspectives, we aim to validate the face validity of the test and ensure that it meets the needs and expectations of its intended purposes. Utilizing a mixed-methods approach, this research combines quantitative data from surveys with qualitative insights from semi-structured interviews. The survey captures broad feedback on test taker experiences, preparation activities, and attitudes towards AI scoring. The interviews provide deeper insights into the new question types and elaborate on the survey findings.

By incorporating the voices of test takers, this research aimed to enhance the validity and reliability of the enhanced PTE Academic test, ensuring it remains a robust and trustworthy tool for assessing English language proficiency.

Methodology

Data collection for this research comprised two methods: surveys and semi-structured interviews.

A survey instrument was built to capture high level feedback on test taker experiences of the test, test preparation activities and attitudes to the use of AI in high stakes testing. All participants who had successfully completed the enhanced PTE Academic and IELTS Academic concordance study tests were invited to complete the survey.

This survey was then followed up with in-depth test taker interviews to gather more detailed feedback, specifically on the two new question types, as well as to provide further elaboration on the topics covered in the surveys. Participants were selected from their survey responses and whether they had indicated that they would like to take part in follow-up research. Efforts were made to ensure that participants invited to the interview represented a range of opinions (from their survey responses) regarding perceived relevancy of task types and skills. Interviews were conducted online via Microsoft Teams. The sample sizes of each qualitative method are shown below:

Research methods used and corresponding sample size	
Research Method	Sample Size
Test Taker survey	663
Test Taker semi-structured interviews	20

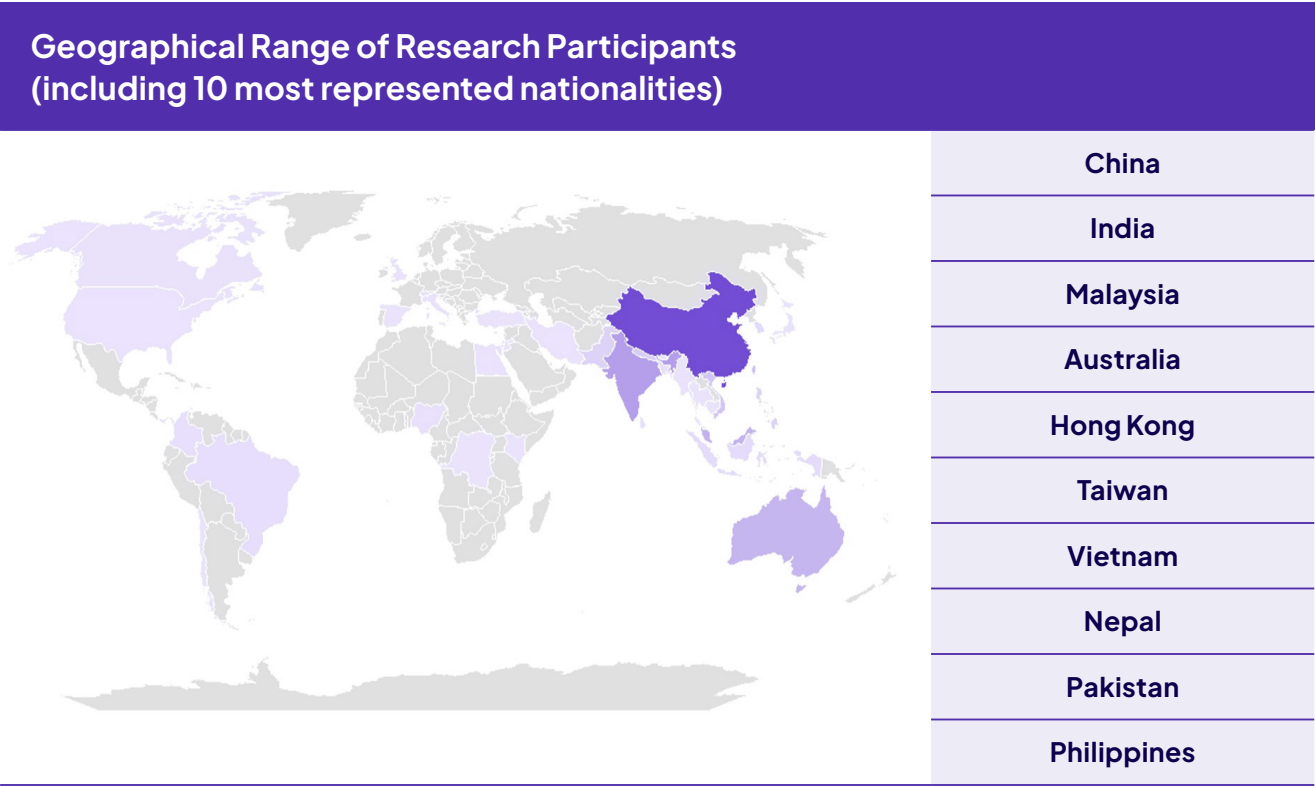
Thematic analysis was performed on the data to establish emergent themes from surveys and then the interviews. These themes are explored in detail in the findings.



Findings

Test Taker Background

Survey data revealed a diverse test taker background, with participants representing 40 different nationalities and speaking 55 different languages. The map below illustrates the geographical range of research participants, with the table adjacent indicating the 10 most represented nationalities. These are representative of the PTE Academic test taking population.



The most common self-reported reasons for wanting to take the PTE Academic test were to facilitate study (36%) and permanent residency/settlement (34%). Interviews then allowed us to probe the backgrounds and motivations of test takers in more detail. Specifically, we were interested in finding out why test takers wanted to migrate to Australia, as well as their long-term goals and aspirations for their lives once settled.

Interviews revealed a wide range of test-taker professions including Healthcare, Accounting, Biomedical Sciences, Web Development and Engineering. Many were pursuing further academic studies, having completed diplomas, bachelor’s or post-graduate degrees. Others were studying for professional qualifications across various industries, such as Heath Administration and Hospitality.

Several test takers mentioned that Australia offered better living conditions compared to their home countries. This included a higher quality of life, better economic opportunities, and faster processes for obtaining permanent residency. One participant discussed this, referencing their preference for Australia compared to Canada and New Zealand:

“Comparatively to Canada, it’s much better. And yeah, compared to the end as well, it’s better I can say. I mean the population [and] the economy wise. And the timeframe it takes for you to get permanent resident here. It’s all better. We had better options, and the universities were good as well.”

(TT6 Interview)

Australia’s education system was a significant draw for some test takers. They appreciated the hands-on and integrated approach to education, which they found more effective than the theory-focused education systems in their home countries. One test taker from India emphasized the differences in education systems between India and Australia, noting the hands-on and integrated approach in Australia:

“It’s like education in India is so saturated because it’s just a competitive place now where there’s hardly any learning done... here it was so hands-on, and everything was so integrated.”

(TT2 Interview)

The ability to work and study simultaneously was a significant factor for many test takers. They found it challenging to balance work and study in their home countries, but Australia provided the flexibility to do both:

“I choose Australia because I think it’s the same idea that everybody abroad has about Australia. It’s a country where you can work and study at the same time.”

(TT1 Interview)

This participant expanded on the benefits of moving to an English-speaking country and that it would increase their opportunities to practice and build on their language learning skills in day-to-day life:

“I am working in a supermarket [while studying]. I have contact with my workmates every day. They are from different parts of the world... we always speak in English because I don’t have any Spanish speaker there... If I stayed in my country... it’s more difficult because there I don’t. There is no, there are not English speakers, so it’s quite difficult to talk or interact with English speakers.”

(TT1 Interview)

Finally, Australia's multicultural environment was appealing to many test takers, as they felt it provided a sense of inclusivity and reduced the likelihood of discrimination:

"I prefer Australia because it's also a multicultural country so...I feel that there will be a limited bias [or] discrimination."

(TT14 Interview)

These reasons highlight a range of factors that make Australia an attractive destination for test takers, including better living conditions, educational and economic opportunities, and cultural diversity.

"The main reason would be the cultural indifferences between people living here. I mean, there are people from several countries, like probably most of the countries of the world, and we are here like it's a cultural mixing. I can be with my friends from different nations from either from America or Europe or Asia itself."

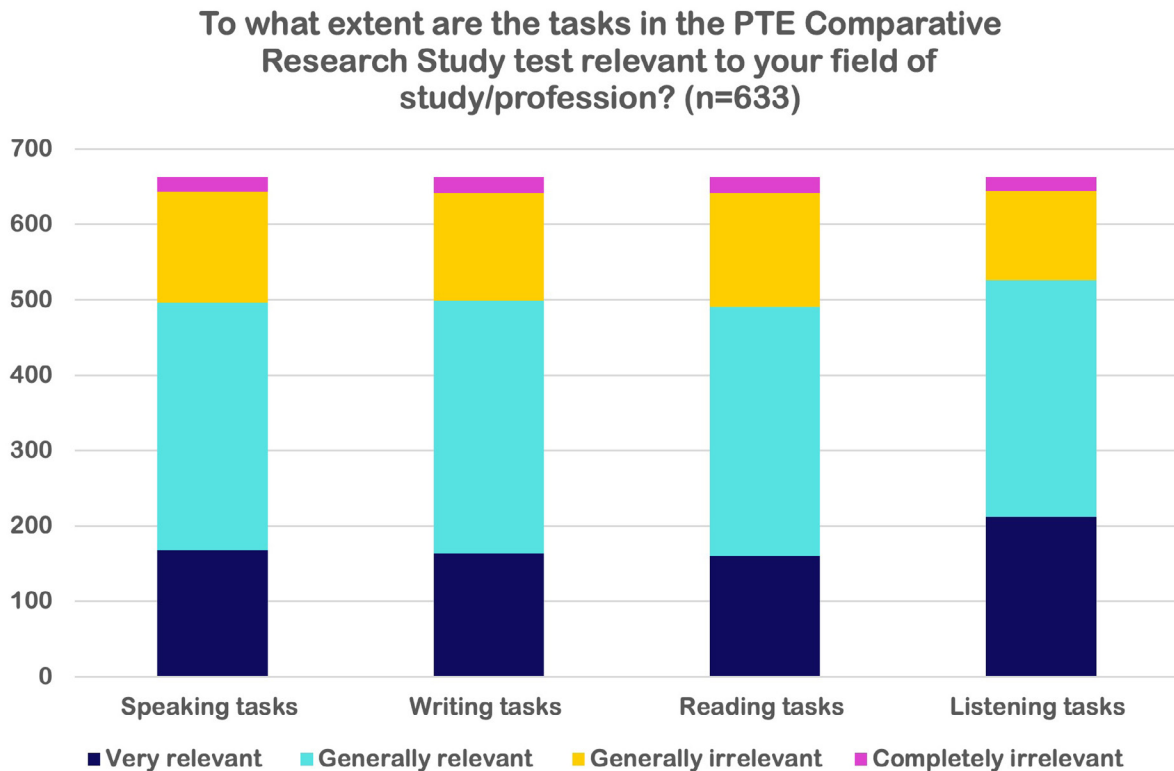
(TT5 Interview)



Test and Question Level Feedback

Feedback on the PTE Academic test was a central tenet to our test taker research at a pre-launch phase. Given the intended purpose of the test, it was necessary to investigate whether task types were reflective of authentic English language as used by test takers in personal, academic or professional contexts.

Chart 1. Perceived relevancy of PTE task types to field of study/profession



Survey data revealed encouraging results regarding test taker perspectives on the PTE Academic task types. Across all task types in PTE Academic, 76% of all survey participants said that they were either generally relevant or very relevant to their field of study or profession.

Regarding the enhanced PTE Academic test, and particularly the two new question types, we also wanted to investigate the extent to which test takers found the tasks (and the skills they test) authentic to activities they may undertake in their real lives. Our survey data elicited responses on specific skills that can be mapped to both the new question types and the perceived importance of these skills to test takers' daily lives. The results are shown in Charts 2 and 3.

Chart 2. Perceived importance of skills tested in Respond to a Situation question type

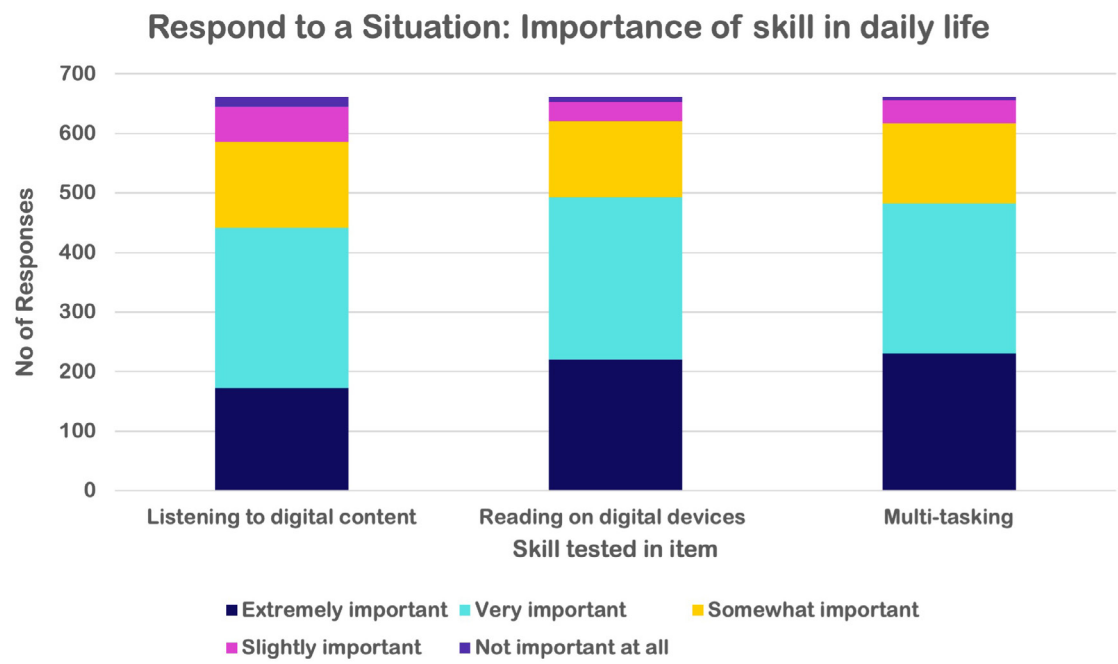
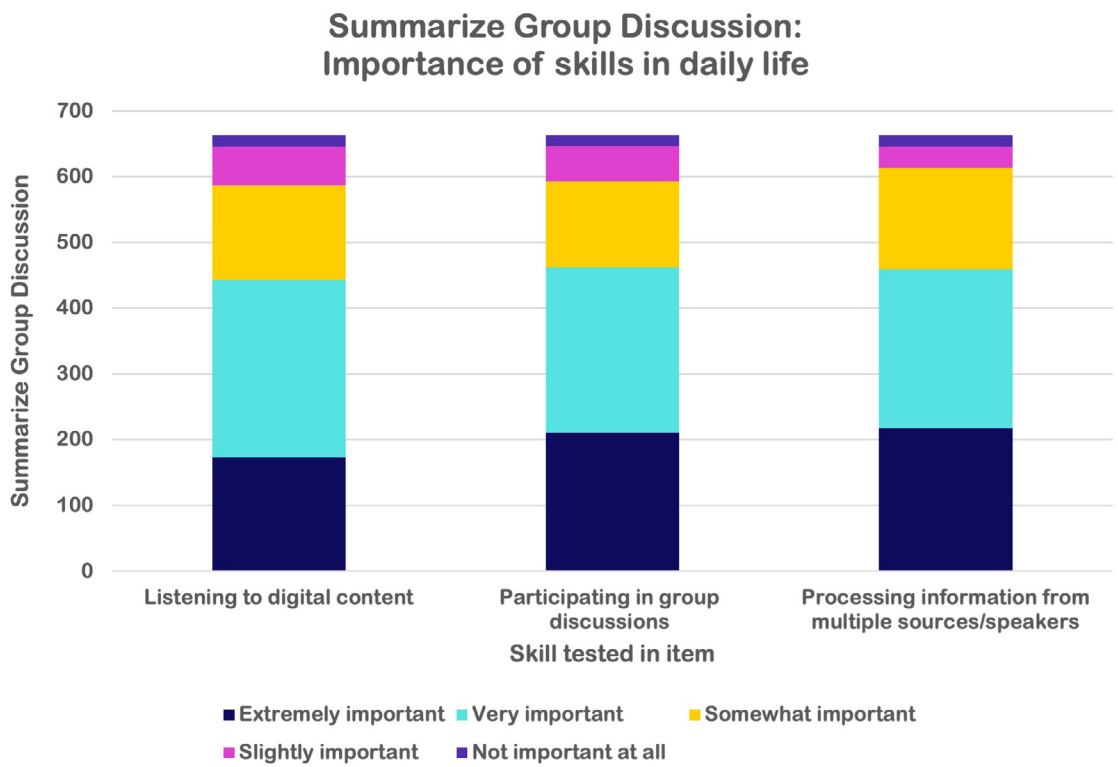


Chart 3. Perceived importance of skills tested in Summarize Group Discussion question type



As illustrated in both charts, 71% of participants felt that Respond to a Situation skills were extremely/very important in their daily life with 69% stating that Summarize Group Discussion skills were extremely/very important. These findings demonstrate that the skills that are being tested are considered important by test takers and reflective of activities undertaken in their daily lives.

This type of evidence plays a valuable and important role in evaluating the face validity of the test, as we ask the question: Do test takers think that the enhanced PTE Academic test is assessing the intended design intentions of the test? As these results highlight, test takers generally think that it is.

Feedback on New Question Types

We sought to obtain more detailed feedback regarding the two new question types. Most notably, we wanted to ascertain whether test takers felt the items were authentic to real-life tasks or situations they encounter in real life.

Respond to a Situation

Most test takers found the Respond to a Situation task to be highly relevant and practical, effectively mirroring real-life scenarios:

"I will say it's extremely relevant... It's a conversation that may come up in daily life, yes."

(TT2 Interview)

"The few questions with this type, I believe it's all very realistic. Very much like college and what typical life will [be like]."

(TT12 Interview)

“

This evidence plays a valuable and important role in evaluating the face validity of the test, as we ask the question: Do test takers think that the revised PTE Academic test is assessing the intended design intentions of the test? As our results show, test takers generally think that it is.

"I think the skill and the content are both relevant to real life. They are pretty realistic and it's not hard to answer like to give an appropriate answer to those questions... It's just like the normal questions that you'll come across in your day-to-day life like someone asking what you are thinking about, to give a suggestion for example."

(TT15 Interview)

Participants appreciated that the task involved giving advice or making decisions, which are common in everyday life. For example, the following quotations are real-life examples given by test takers where these skills have an authentic application:

"This scenario can happen in your real life. That's one thing that I why I feel this is fair. It was practical. You open see someone leaving your job last day of your job or someone's last day at your job and then you have to discuss with your friends. OK, then we have to plan this for this particular person. So it was relevant."

(TT10 Interview)

"Maybe you are at the intersection on a main highway, and you are taking right turn. Your friend is suggesting not to take right turn, but you are thinking that you should take the right turn."

(TT5 Interview)

"In [a] restaurant in the kitchen... There are situations where you have to come up with something else... A lot of times people come with... allergies, so we have to come up with a different recipes for them and, you know, thinking on your feet like there are a lot of things, there are a lot of scenarios."

(TT19 Interview)

Another test taker emphasized the significance of quick thinking and noted that this skill was frequently practiced during study sessions:

"Thinking rapidly is what I came across while studying as well. We had a quick-fire round of information exchange, like two or three students are sitting together, someone asks you something and you need to respond in like one minute or two minute."

(TT5 Interview)

The task was seen as relevant for forming significant relationships, whether at university, work, or in social contexts. Test takers emphasized the importance of social connections, especially when moving to a new country without family:

“I feel because it has a lot of social aspect to it, which I think is important as well because you’re not just going to a place to just study, you’re obviously going to live there.”

(TT2 Interview)

To this extent, the Respond to a Situation task was considered to be applicable to both professional/study contexts as well as everyday situations. One test taker described the task as “genuine”, reiterating the importance of:

“actually listening to someone, actually paying attention to what they’re saying and trying to get to... the points that they want to go to.”

(TT18 Interview)

Generally, most test takers noted that it’s similarity to everyday interactions made the task quite straightforward to answer. Others found it challenging due to the need to think quickly and manage the response time effectively. They mentioned that the question tests their ability to think quickly and respond appropriately, which can be demanding.

Some participants found the short preparation and response times to be difficult, as it required them to formulate their answers rapidly:

“I think that’s probably the hardest part of this, yeah. Is knowing, yeah, knowing just the right amount to say.”

(TT17 Interview)

It is worth noting, however, that the demands mentioned by test takers regarding this task do not necessarily indicate issues with the task itself but instead reflects better opportunity to differentiate between English language proficiency levels.

“ The Respond to a Situation task was seen as relevant for forming significant relationships, whether at university, work, or in social contexts. Test takers emphasized the importance of social connections, especially when moving to a new country without family.

Summarize Group Discussion

Many test takers found the Summarize Group Discussion question type to be highly relevant and realistic to real-life scenarios. Feedback mentioned that it mirrors professional and academic settings where summarizing discussions and understanding multiple perspectives are common tasks.

As one test taker explained:

"I think it is pretty much relevant to like what the English tests should be. Because in day-to-day life we come across such situation like that people are talking like for an example, like I'm working and... I have two colleagues; we are sharing some important information regarding our work."

(TT5 Interview)

Several test takers gave specific examples of where they have experienced a similar scenario to the Summarize Group Discussion task in both work and study contexts. The following three quotations provide examples of where test takers felt the Summarize Group Discussion task was relevant to their real lives:

"I could say that in professional life, I think it would be really good because [it] doesn't matter wherever your career is, maybe if you are a teacher... if you have [a] meeting with your other peers. Sometimes... a secretary maybe [has] to summarize all the meeting all the things that you hear there and after that send to the people who was in the meeting, maybe with the key points that you need to show them in order maybe to make an agreement or maybe to do whatever thing that the company needs."

(TT1 Interview)

"Sometimes we have discussions, or I have to listen some discussions related with the health problems of our patients. For example, if is a paediatric patient, we have mom and dad in the same room...and we have to pay attention what they are discussing or what are the key points or maybe the main problem or the main health problem of the patient, so I think it's relevant."

(TT1 Interview)

"In my second year I joined a few student clubs and sometimes a lot of university professors are also part of those clubs. And yeah, we do have a lot of discussions on what could be good for like open days like when we set up our stalls and what wouldn't work out as well. And because I was there in that club for like 3 years. So over the years I saw how different things work better."

(TT2 Interview)

Test takers also saw the benefits of this question type to daily life tasks, such as participating in group discussions socially:

“For example, when I live with my housemates, we usually speak in a group and share our thoughts over something. It will be a scenario kind of like this.”

(TT20 Interview)

“I would imagine this would be like a normal conversation we had with my friends.”

(TT3 Interview)

This was reiterated by another participant who discussed that the skills assessed by Summarize Group Discussion are important in facilitating connection with different people:

“What you want to say, [it’s important] regarding...your study or whether you just want to interact and socialize because whatever you need, you have to discuss, or you have to convey, or you have to understand other people’s interaction.”

(TT4 Interview)

Another participant described taking part in group discussions while doing their Master’s degree, stating that English language proficiency was crucial to communication:

“It’s really important when you are talking to a group of people, [a] group of people maybe [has] people from different regions of the world, they might have [a] different accent as well. But the only thing common in a group discussion would be the language that we are communicating in. And most probably it’s English.”

(TT5 Interview)

This comment is a useful reminder of the ubiquity of English language skills in regard to the integration in distinct cultures, communities and first languages, emphasizing how plurilingualism operates in many academic and professional communities.

Feedback suggested that test takers found Summarize Group Discussion challenging due to its length, complexity, and the need to summarize multiple speakers:

"I think in my opinion, the topic itself, it's quite relevant, but I think the time frame is a bit hard because...if we are conversing together, we only like we only need to memorize one or two word, one or two sentences and then we will make a reaction."

(TT12 Interview)

Test takers highlighted the difficulty of maintaining fluency and memory retention under time constraints:

"The requirement of fluency is quite high, as it requires you to have less pause as possible between sentences."

(TT15 Interview)

Some participants noted that the task was more demanding compared to other questions, such as Respond to a Situation, due to the quick pace and the need to retain and process a lot of information:

"This probably was a bit more challenging than the worded ones. You can, you know, read them again...But if you missed something [in Summarize Group Discussion Question]... if you miss it, you know, it's gone. You can't go back and play it."

(TT17 Interview)

The fact that some test takers found this question type challenging is not necessarily detrimental to the validity of the intended construct. This question provides another means of better evaluating proficiency levels and providing opportunities to demonstrate a broader range of skills.

Test Preparation

Test takers use a wide range of materials and resources to prepare for PTE Academic, so we also asked them about their test preparation practices. It is worth noting, however, that the circumstances in which test takers sat the enhanced PTE Academic test at this stage of development differed to usual circumstances. The enhanced test was in a research phase at the time of testing and did not have currency in terms of applications to universities or professional bodies. We made efforts to address this potential limitation in our interview questions, by also asking test takers to consider preparation practices they may have used when sitting a current PTE Academic test.

Our survey data showed that test takers used a combination of different resources to prepare for PTE Academic. While free preparation materials were the most commonly used resource (34%), these were often used in conjunction with social media groups, paid preparation resources, the PTE offline practice test and materials available on the PTE website. We were able to discuss this in more detail during test taker interviews, with some of the key practices outlined further below.

Many test takers used various online applications and apps to prepare for the test. One test taker mentioned:

"I had the PTE [Academic Practice] app on my phone for the practice test and I tried it and I thought it was quite easy cause a lot of people could even comment the answers as well."

(TT2 Interview)

Another test taker shared their experience of using a third party app to help prepare:

"One of my workmates showed me an app... I put my target as 65 plus... I paid 30 Australian dollars for one month and they gave me tasks every single day."

(TT11 Interview)

Mock tests and practice questions were also frequently used by test takers. One participant mentioned:

"I took some free mock tests which the PTE website offers... I also did paid mock tests... My preparation was related to only mock tests... mostly mock tests only."

(TT9 Interview)

YouTube videos were a very popular resource used among test takers to familiarize themselves with the structure of the test and style of questions. One participant said:

"I initially started seeing the videos which did the practice tests... I like that some of them actually broke down what is expected from the test."

(TT2 Interview)

Another test taker found using YouTube videos in their native language helpful, stating:

"I just randomly go to YouTube and find some tutor... they can talk in Punjabi too."

(TT4 Interview)

Social media groups and online forums provided valuable insights and tips from other test takers. As one participant shared:

"The social media groups... they just discuss what type of questions are coming in PTE these days... when people give their test, whatever they remember, they just pop it up on social media."

(TT6 Interview)

It was evident that Pearson's official materials were used by some test takers to build a strong foundation. One test taker mentioned:

"The materials from Pearson's are really helpful because it's a standard material... it can help the first-time tester to understand and follow the structure of the test."

(TT14 Interview)

They continued, stating:

"For new test takers, Pearson's materials are a strong source, and for experienced test takers, online resources and forums can give useful tips and tricks."

(TT14 Interview)

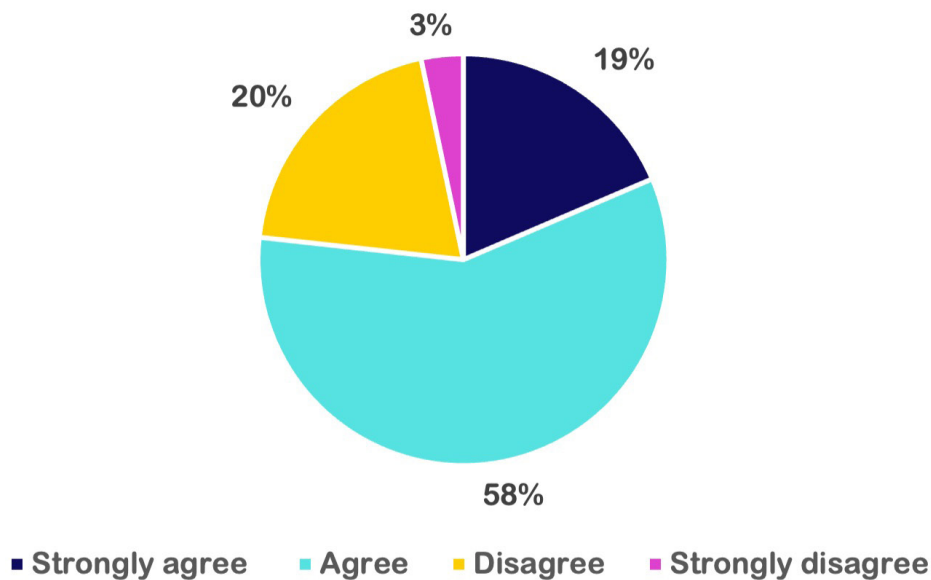
Attitudes to AI Scoring in High–Stakes Tests

With growing interest in and application of AI in education and employment, we were interested in capturing perspectives on its use in scoring of high stakes language testing. Our test taker sample were familiar with AI tools (e.g. Chat GPT, Copilot), with 58% of survey participants using them regularly for educational, professional, and general purposes. Whilst it was not an intentional focus of this research, we were interested to see if familiarity and use of generative AI tools had unintended consequences on perceptions of the use of AI in test scoring. In both surveys and interviews, we questioned the extent to which test takers believed tests marked by AI are fair, reliable and trustworthy, while also probing notions of objectivity, bias and the relationship between AI and human scoring.

Survey data showed positive feedback regarding the trust test takers have in AI scoring systems and the perceived objectivity in comparison to human markers. As Chart 4 shows, 77% of survey participants agreed/strongly agreed that tests marked by AI are fair, reliable and trustworthy.

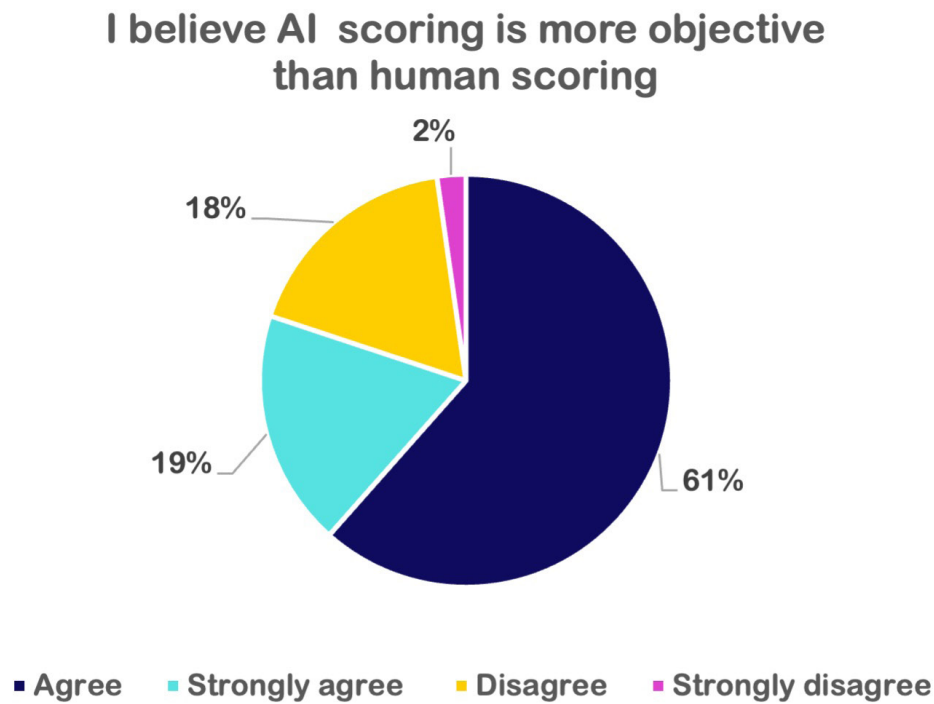
Chart 4. Perceptions of fairness, reliability and trustworthiness of AI scoring

**I believe tests marked by AI are fair,
reliable and trustworthy (n=633)**



Survey data also gave optimistic results regarding perceptions of objectivity in AI scoring, as 81% of participants agreed/strongly agreed that AI scoring is more objective than human scoring.

Chart 5. Perceptions of objectivity of AI scoring compared to human scoring



Once again, we were able to capture more nuanced feedback in test taker interviews where participants discussed topics of fairness and objectivity in AI scoring.

A common thread in test taker interviews centered on the objectivity in computer-based testing and AI scoring, an aspect that increased perceptions of trust. Many test takers discussed concerns regarding biases with human interaction, something alleviated by computer-based testing. For example, several test takers noted that completing their test solely on a computer significantly reduced pressure and anxiety compared to interacting with a human assessor:

“Whenever I have to interact with the actual person, then there will be [a] different kind of pressure. Sometime is because of the other person’s attitude or kind of like looking [a certain] way at us kind of thing makes me nervous, which can be distracted from my actual performance.”

(TT4 Interview)

This test taker continued, explaining how the structure of PTE Academic (i.e. machine scored) alleviates their nervousness and reduces the feeling of being judged:

“In that way, I just find [the PTE test] so, so comfortable...I’m just very calm and relaxed the whole test because I just completely rely on the machine and just doing my result and... I’m just confidently speaking without any pressure or anything like that... Sometimes there are situations [when interacting with a human] when you just go out of track, then the person maybe just makes a different kind of expression which make you like more even nervous. With a machine, I just don’t feel like that kind of thing even though sometimes I know...that some of my questions may be wrong, but still there will be no nervousness or anything because I know there’s no actual person who can judge me.”

(TT4 Interview)

Participants also highlighted the fairness of AI scoring systems, particularly in comparison to human scoring. One test taker stated:

“I believe there’s less chances of biasedness when it comes to AI testing. You will get biasedness in somewhat of human testing as well.”

(TT5 Interview)

This was supported by another participant who appreciated that:

“AI scoring will not affect my results by emotions or personal perspectives.”

(TT14 Interview)

Although data indicates trust in AI scoring systems regarding fairness and objectivity, feedback from test taker interviews relayed a sense of pragmatic scepticism and cautious optimism, with some having doubts about its accuracy.

For instance, while one participant acknowledged the efficiency of AI scoring systems, they expressed some reservations about the complete reliance on AI to mark tests:

“I believe the models that are for training AI...are much more efficient. But I would say human involvement in any kind of test needs to be there because AI can make mistakes.”

(TT5 Interview)

This comment reflects a general attitude amongst many test takers, whereby they expressed support for the use of AI with an element of human involvement. The following quotations illustrate this, mentioning that human involvement with marking would bring a peace of mind:

"I would prefer like a combination of AI and human marking... the human marking can be like a safety net if the AI doesn't work well."

(TT14 Interview)

"Having a human element, I think would really especially... bring another level of confidence."

(TT18 Interview)

Other test takers also mentioned that human involvement in scoring would be beneficial for speaking tasks specifically:

"In terms of reading, writing and listening part, I think the computer is a lot fairer than the human. But when it comes to speaking, I think it'll be a lot better if a human were to judge compared to AI."

(TT19 Interview)

"Definitely a mix. So, I feel like anything to do with writing can easily be, you know, run using AI 'cause after all, they are just words and there are no changes to words, whereas with things that are spoken because sometimes accents and such. A human would probably be more adept at listening out to things that you have to say."

(TT8 Interview)

Conclusion

This research provided valuable insights into the perspectives of test takers regarding the revised PTE Academic test during the phases of research and development.

The findings highlighted the diverse backgrounds and motivations of test takers, emphasizing the importance of understanding their experiences to enhance the validity and reliability of the test. The feedback on the new question types, Respond to a Situation and Summarize Group Discussion, indicated that these tasks were perceived as relevant and reflective of real-life scenarios, thereby supporting the test's face validity.

Moreover, the study reveals a general trust in AI scoring systems, with test takers appreciating the objectivity and fairness they offer compared to only human scoring. However, there is a consensus on the need for a balanced approach, combining AI and human oversight to ensure accuracy and reliability, particularly for speaking tasks.

Overall, the research underscores the significance of incorporating test taker feedback in the test development process. Through a combination of empirical data from trialling, field testing, and concordance phases of development, as well as input from various stakeholders, the validity and reliability of the enhanced PTE Academic test has been established. Both quantitative and qualitative stands of validation evidence will continue after the launch of the enhanced PTE Academic test, ensuring it remains a robust, trustworthy and fair assessment that meets the needs of both the test takers and the test accepting communities.



About the author



Ellen Barrow

Ellen is a Senior Researcher within the PTE Assessment Research and Validity team, undertaking research studies on a range of PTE test validity related areas with a particular focus on test-taker experiences. Ellen has previous experience in the design and development of a range of onscreen assessments and has led impact evaluation research as well as supporting the delivery and evaluation of the global survey instruments of PIRLS, TIMSS and PISA in England. Key aspects of her work have concentrated on the equity and fairness of assessments from a user-centred perspective, aiming to improve accessibility of digital assessments, remove barriers, and enhance the learner experience. Ellen is an accredited member of the Association for Educational Assessment – Europe (AEA-E). She received her BA from the University of Reading and her MA from Oxford University.



Secure. Accurate. Trusted.

pearsonpte.com/research