



Full Report

**Examining the relevance of
PTE Academic for Australian
professional bodies**

Contents

Introduction	3
Background	3
Stakeholder panel	7
Method	9
Activity 1 – Relevance of tasks	9
Activity 2 – Level of proficiency	11
Findings	12
Activity 1 – Relevance of tasks	12
Speaking	12
Writing	16
Reading	17
Listening	19
Activity 2 Findings – Level of proficiency	19
Performance descriptors (draft)	22
Superior	22
Proficient	22
Competent	22
Vocational	23
Conclusions and future research	26
Acknowledgements	28
References	29
Appendix A – Worksheet Exemplar	30

Introduction

This paper reports on the findings from a Test Review Group (TRG) who were asked to investigate the domain relevance of PTE Academic to the Australian skilled migration context. **Engaging directly with professional bodies which include English language proficiency as a requirement for membership to their organisations.**

Professional bodies in Australia set their language requirements in reference to the DHA language proficiency categories. Based on the TRG findings, performance descriptors for these categories were drafted to reflect the knowledge, skills, and abilities required at different levels of language proficiency within the context of skilled economic migration.

Background

The PTE Academic test was initially purposed to assess the language proficiency of test takers entering higher education for academic purposes in English speaking countries. From 2014, the test was also accepted as part of the entry requirements for skilled economic migration into Australia. In particular, the test is used primarily for mid to higher-end visa categories and used as a language gatekeeper for a range of post-graduate occupations and professions.

The focus of this research study and paper is to explore the issue of Australian economic migration purpose in regard to the PTE Academic test and the subsequent validation approaches.

Newton (2007) identified three main purposes of an assessment system:

- 1. to measure;**
- 2. to make decisions; and**
- 3. to achieve impacts.**

Each of these three main purposes can be subdivided, depending on the phase of education, regulatory and policy involvement, however essentially each claim requires a different kind of supporting validity argument: a measurement argument; a decision-making argument; or an impact argument. Often, though, an impact argument will subsume a decision-making argument; and a decision-making argument will subsume a measurement argument.

The issue of repurposing tests has always been contentious in, and in the context of English proficiency tests it has become a source of debate amongst university language centres in Australia. Where a test has been developed and validated for a particular purpose, it can be said to have a mandate for test use in a given context. However, “mandate creep” often occurs when societal and organisational forces require an existing assessment to be used beyond its original purpose (Macqueen et al., 2021). In this situation, assessment organisations have a responsibility to monitor and evaluate the evolving use of tests to ensure that score interpretations remain valid.



Although there are many different conceptions of validity, the most unifying consensus definition used by international testing organisations is taken from the *Standards for Educational and Psychological Testing*:

This definition recognises that test results can be used for multiple purposes and a test fit for one purpose may not be fit for another. As such, the idea that a test instrument can be declared either valid or invalid is inappropriate. Instead, since results from a single test can be interpreted in different ways for different purposes, it is more appropriate to refer to the validity of a particular interpretation of test results.

As part of our responsibility to monitor and evaluate PTE Academic score interpretations, Pearson has previously awarded research grants to the University of Melbourne Language Testing Research Centre to investigate the evolving use of PTE Academic in Australian skilled migration pathways, including nursing and accounting professions (Frost et al., 2018; Knoch & Pitman, 2019). These studies investigated stakeholder perceptions of the relevance of PTE Academic by interviewing former test takers who had recently achieved qualifying scores on PTE Academic and entered the workforce in their profession.

Both studies reported partial support for the relevance of PTE Academic to two professional fields, and recommended further research in this area.

Both studies focused on test taker perceptions. Although the sample sizes were small and opportunistic in terms of participants, they offered useful insights of language tests to new members of two professions, at a point of beginning to develop a sense of the standards of linguistic proficiency required to successfully navigate a range of professional duties.

Most approaches to validation studies build on Kane's (2001, 2006, 2013) conceptualization of inferences, warrants, and assumptions, thinking of a validation argument as a chain of reasoning that generalise from performance in an assessment situation to performance in the real world. The key inferences are the:

- scoring inference – the test score makes a claim on the basis of the observed performance;
- generalisation inference – the universe score makes a claim on the basis of the test score
- extrapolation inference – the score interpretation makes a claim on the basis of the universal score.

In collecting validity evidence it is therefore essential to include as many relevant contributions as possible. In the case of this research study, the focus is to expand the stakeholder voice in the context of informed perception of the fitness for purpose of the PTE-Academic test for professional body entry via DHA visa categories.

An example of an operational validation framework for use in high stakes assessments is given by Shaw and Crisp (2012; 2015). Like Kane, they use an Interpretation and Use argument, and

conclude with a decision-making inference. Unlike Kane, they begin with a construct representation inference. It is important to note that Kane's conceptual framework does not insist upon any particular argument structure. He just recommends that the argument should be coherent, that its inferences are reasonable, and that its assumptions should be plausible.

The table below is taken from Shaw and Crisp (2015, p.36, Figure 3) to illustrate how this kind of framework can be used to identify key questions for validity research.

Inference	Warrant justifying the inference	Validation questions
Construct representation	Tasks elicit performances that represent the intended constructs	1. Do the tasks elicit performances that reflect the intended constructs?
Scoring	Scores/grades reflect the quality of performances on the assessment tasks	2. Are the scores/grades dependable measures of the intended constructs?
Generalisation	Scores/grades reflect likely performance on all possible relevant tasks	3. Do the tasks adequately sample the constructs that are set out as important within the syllabus?
Extrapolation	Scores/grades give an indication of likely wider performance	4. Do the constructs sampled give an indication of broader competence within and beyond the subject?
Decision-making	Scores/grades give an indication of likely success in further study or employment	5. Do scores/grades give an indication of success in further study or employment such that they can be used to make appropriate decisions?

“Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests.”

(AERA, APA and NCME, 2014, p.11)

The research questions that this study set out to investigate were:

- Are PTE Academic test items relevant and valid for assessment of English language proficiency in the domain of professional body visa entry requirements in Australia?
- Are the skills elicited by the PTE Academic item types relevant and valid within professional bodies in Australia?
- Do PTE Academic performance standards as categorized by Australian visa categories match professional bodies expectations of English Language proficiency?
- Can Performance Standards be developed to identify the key competencies elicited in English language proficiency tests at each visa classification in Australia?

The research study reported on in this paper investigates these research questions in the context of the fitness for purpose of the PTE Academic test to skilled migration pathways by engaging with stakeholders who are responsible for the standards of proficiency required by their field and who are well acquainted with the diversity of roles and skills within their professional bodies in Australia. It was important to create an evidence gathering environment where the validity questions were clearly understood to professional body practitioners with little research backgrounds. Therefore the table above illustrates how the assessment use questions can be cross referenced to an interpretive validity argument.

In exploring the relevance and validity of PTE Academic, this study also seeks to better describe the proficiency requirements of professional bodies in relation to the DHA language proficiency categories. Professional bodies in Australia set their language requirements in reference to the DHA language proficiency categories: Functional, Vocational, Competent, Proficient and Superior. Unusually, performance descriptors of these categories do not exist.

The DHA language proficiency categories were originally set in relation to IELTS test score bands, and as such do not include skills-based holistic performance descriptions. As other testing agencies have been accepted by the DHA, equivalence of scoring categories have been used, therefore there is no public information regarding what these proficiency categories look like in practice, particularly in relation to the professional and occupational usage of English language. As such, before it is possible to fully evidence the relevance of PTE Academic to professional bodies' language requirements, these requirements must be better understood beyond their relationship to IELTS score bands and these requirements must be articulated as performance descriptors.

Performance descriptors are holistic statements that define the knowledge, skills and processes of test takers at specified levels of achievement (Egan et al., 2012). They are usually created by a combination of inputs from policy makers, stakeholders and content experts in order to create a common understanding of required performance standards at specified levels of achievement. The primary purpose of language proficiency tests is not to determine the rank ordering of test takers, as is the case with norm-referenced tests, but rather to determine the placement of test takers into a set or ordered categories (Hambleton, 2012).

Therefore, performance descriptors for these categories are purposed to ensure that standards are understood and maintained when they are set or when multiple test providers are involved in awarding equivalent qualifications. They also provide a straightforward description of what particular levels of achievement look like in practice and therefore aid understanding and transparency for test takers, accepting institutions and policy makers.

As one of the test providers, Pearson does not own the DHA proficiency standards. However, as assessment experts, Pearson is able to carry out robust standards-related activities and procedures. This paper reports on research to work with Australian professional bodies to explore their language proficiency requirements and their relationship to both the PTE Academic test and the DHA language proficiency categories. Through this investigation, working alongside professional organisations, performance descriptors were developed for DHA proficiency categories that are rooted in the requirements of professional bodies and based on test taker and test item exemplification across the four skills of language proficiency.

The emergent performance descriptors are aligned to the PTE Academic test, however, as they are holistic statements, they provide a working description of performance standards for DHA proficiency categories that should be applicable across test providers.

Test Review Group (TRG) Stakeholder panels

The TRG stakeholder panels were composed of twelve representatives from nine professional bodies that currently recognise PTE Academic as evidence of English language proficiency in Australia. Representatives were nominated from the senior leadership of their organisational groups. They were all experienced educators and practitioners who have had involvement over many years in selection processes into their professions. They all had an interest and understanding of entry assessments and language requirements and also how and why English language proficiency is essential to maintain the quality and safety of delivery in their sectors.

Invitations were sent to a range of professional bodies in Australia. The response to the invitations was positive, however due to scheduling issues, the TRG was split into two sessions. The first meeting was conducted in person in Melbourne, the second meeting was conducted on-line. For the in-person event in Melbourne, five professional groups attended:

- Australian Nursing & Midwifery Accreditation Council (ANMAC)
- National Accreditation Authority for Translators and Interpreters Ltd (NAATI)
- Engineers Australia (EA)
- Australian Computer Society (ACS)
- Australian Institute of Medical and Clinical Scientists (AIMS)



The second online meeting was attended by:

- Certified Public Accountants (CPA)
- Chartered Accountants (ANZ)
- The Australian Society of Medical Imaging and Radiation Therapy (ASMIRT)
- Australian Pharmacy Council

Panellists completed a survey detailing their professional experience. **The majority of the panel had more than 15 years of experience in their professional field, with all participants having at least 5 to 9 years**

of experience. All participants were familiar with English language assessment, the language proficiency requirements of their organisation, and the DHA language proficiency categories.

All professional bodies represented by the two panels use some combination of Competent and Proficient language requirements for accreditation purposes. The language proficiency requirements of PTE Academic scores across language skills for these professional bodies are summarised below:

Organisation	Language proficiency requirement	PTE Academic score				
		Overall	Listening	Reading	Speaking	Writing
NMBA	Proficient	65	65	65	65	65
EA	Competent		50	50	50	50
AIMS	Proficient	65				
NAATI – Recognised Practising Interpreter and Certified Provisional Interpreter	Competent		51		59	
NAATI – Certified Interpreter and Certified Specialist Interpreter	Proficient		76	76	76	59
NAATI – Certified Conference Interpreter	Proficient		76	76	76	59
NAATI – Recognised Practising Translator	Competent			59		59
NAATI – Certified Translator	Proficient			76		76
NAATI – Certified Advanced Translator	Competent		59	76	59	76
CA ANZ	Proficient	65	65	65	65	65
CPA A (general skills assessment)	Proficient	65	65	65	65	65
CPA A (provisional skills assessment)	Competent	50	50	50	50	50
ASMIRT	Proficient	65	65	65	65	65
APC	Proficient	65	65	65	65	65

Method

The aim of the Test Review Group (TRG) activity was to investigate the relevance of PTE Academic to the standards of language proficiency required by each professional body. Standards were not being set in this activity. The DHA owns the standards for proficiency categories and the PTE Academic cut scores have already been set in respect to these standards. However, the activity aimed to investigate and describe standards, and so elements of the activities were based in best practice approaches to standard setting. We aimed to incorporate the Council of Europe's (2009) guidance on the appropriate steps in standard setting into aspects of the activity, including:

- Familiarisation
- Specification
- Standardisation training/benchmarking
- Standard setting
- Validation

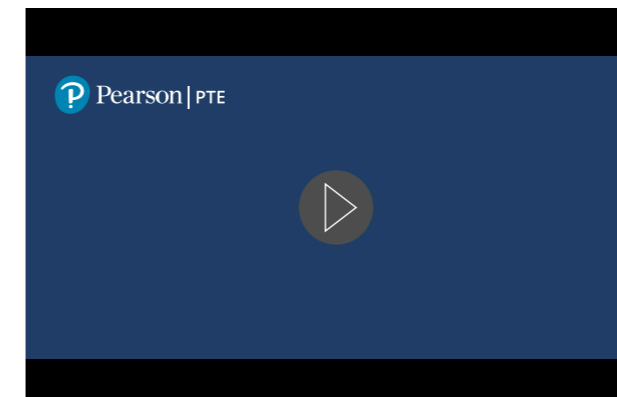
For the activity, panellists were asked to keep in mind the potential job roles of someone newly qualified in their field.

The purpose of this was to keep the judgements of each panellist consistently attuned to the way in which language proficiency requirements are used at the threshold of entrance into their own organisation. Relevance was investigated both in terms of the skills required to complete assessment tasks and in terms of the qualities of test taker responses that indicated proficiency.

Activity 1 – Relevance of tasks

To effectively comment on the relevance of PTE Academic, it was necessary that panellists should be familiar with both the PTE Academic test and the language proficiency requirements of their professional bodies. The Panellists were all assessment specialists within their fields, and therefore were sufficiently familiar with their associated language requirements. To ensure familiarity with the PTE Academic test, panellists were provided with information about each PTE Academic item type, including the response requirements, the scoring criteria, and the skills intended to be elicited by the item type. An example of the information provided for each item type is shown below.

Read aloud



- Test takers read a brief passage out loud.
- Test takers have 30–40 seconds to prepare before reading the passage aloud.
- Scored on content, fluency, and pronunciation.
- This item type is designed to assess test takers' ability to read the content of a passage accurately, and to communicate the content clearly.

Panellists were also shown the Sample Assessment Materials (SAMs), which are composed of exemplar items representing a complete test form. The SAM is a sample test comprised of items that have been removed from live operation, and therefore do not compromise test security. For each of the items in the SAM, Pearson provided item level performance data, and also audio and written test taker exemplars for the speaking and writing items, taken from across the scoring range.

Following the familiarisation briefing, panellists were asked to comment on the relevance of the tasks. For each item type, panellists were asked to complete a worksheet with two open response questions (see Appendix A for an exemplar worksheet). The questions asked them to comment on:

1. Whether the item type is similar to any tasks in their profession.
2. Whether the skills elicited by the item type are important within their profession.

While these two questions may seem similar, they are intended to each have a distinct focus. The first question aims to investigate whether the item type itself is an authentic replica of tasks that are carried out within a profession. The second question aims to understand whether the underlying skills required to answer the item type are important within the professional field. It may be that the answers to both questions are aligned. However, by separating the two questions, panellists are encouraged to also consider instances where the task of completing an assessment item is not a direct replica of a task within their profession, but where the underlying skills assessed by the item are integral to the profession.

As part of the second panel TRG meeting, we included a slightly amended version of the SAM materials, containing two additional item types. These item types have been used in other PTE language proficiency tests and are now being considered for use in PTE Academic in order to add more language complexity and real-life authenticity to the test as a whole.

Activity 2 – Level of proficiency

For speaking and writing item types that require open responses, as well as discussion on the item types themselves, panellists were asked to comment on the level of proficiency demonstrated in actual test taker responses. A ‘body of work’ approach was taken where a number of open responses were presented from each test taker. This meant that rather than an impressionistic view on a single response, the panels were able to formulate a holistic view of the proficiency levels of each test taker.

For each item type, approximately 10 test takers were sampled, with roughly equivalent numbers of test takers being drawn from the DHA proficiency categories: Superior, Proficient, Competent and Vocational. DHA classifications are based on the test taker’s scores across all four communicative skills: Listening, Reading, Speaking, and Writing. As the on-line meeting was shorter, fewer test taker exemplifications were used, although the spectrum of low, medium and high level of scored proficiency was maintained.

The test taker responses were presented to the panel in descending order of classification, beginning with Superior and ending with Vocational. The panel were informed that the sample was ordered by test taker classification. The scores for individual responses were not shared with the panel.

The aim of this activity was to demonstrate the kinds of responses produced by test takers classified at different levels, and to focus on the aspects of their performances that were or were not indicative of a minimally qualified candidate in each professional field. We therefore encouraged a holistic view of the test taker, rather than focusing on the scores to individual responses. The panellists were asked to indicate whether each response contained English language ability that was acceptable from someone who has just qualified in their professional field. These judgements were captured on the same worksheets described above. An example of the worksheet is provided in Appendix A.

Activity 1 followed the order of the item types in the live PTE Academic test beginning with speaking and writing item types, followed by reading and listening. In the second on-line meeting, two additional speaking items were added to the speaking section.

Activity 2 was undertaken following Activity 1 in the speaking and writing sections – the productive skills. Activity two did not include the single skill items of reading and listening as these do not provide student exemplification.



Findings

As described in the method of this paper, it was necessary to familiarize the panels with the underpinning construct of language proficiency testing in general and the form and structure on the PTE academic test specifically.

Most international language proficiency tests are based on the language construct as laid out by the Common European Framework of Reference (CEFR). The panels were aware of the four skills requirements for visa categories and that all major language proficiency tests use a common construct as the basis of their test. The panel were also aware that language proficiency tests in Australia are used for study and economic migration purposes. The difference between tests is in the form, structure and mode of the test.

An overview of the PTE Academic test was described to the panel and then each language skill section was shown and discussed, going through every item type.

Activity 1 – Relevance of tasks

One of the objectives of the TRG was to ascertain whether the item types used in PTE Academic were relevant to the domain of English language proficiency in the context of skilled immigration to Australia.

In the course of the Activity 1, panellists recorded their comments on the relevance of tasks and discussed as a group the ways in which the targeted skills manifest in job roles within their field. There was broad consensus across the panel that PTE Academic items have relevance to professional tasks undertaken in range of fields. The panel's comments are summarised below for each section of the test¹.

Speaking

Across the Speaking items, there were tasks or underlying skills that were relevant to all the professions represented on the panel. Speaking item types include:

- Read Aloud
- Repeat Sentence
- Describe Image
- Retell Lecture
- Answer Short Question
- Summarise Group Discussion²
- Responding to a Situation³

For the Read Aloud and Repeat Sentence items, item exemplification emphasised that these items did not entail paraphrasing or summary skills. Those skills are assessed in other item types. Read aloud and repeat sentence test accurate recall and also assess effective pronunciation and fluency. Panellists focused on professional settings in which information needs to be read aloud or repeated verbatim to ensure safety and adherence to regulatory protocol. Panellists offered a variety of examples of professional tasks that require this type of precise verbal communication. Examples are given here, that are representative across the two panels.

- “Yes – reading out methods when training staff, communication of test results using pro-forma sentence.” (AIMS1)
- “Most situations would involve reporting test results clearly and accurately to clinical staff.” (AIMS2)
- “Yes – reading, interpreting and speaking it aloud is a skill required in the nursing and midwifery professions, e.g. reading out procedure to draw up drug infusion in a high pressure/high stakes environment (ICU/ITU).” (ANMAC1)
- “Instructions on procedures and processes, e.g, phlebotomist interacting with patients – informed consent. Training new staff on techniques and procedures etc.” (AIMS2)

- “Accountants provide services relating to financial reporting, taxation, auditing, insolvency, accounting information systems, budgeting, cost management, planning and decision-making by organisations and individuals; and provide advice on associated compliance and performance requirements to ensure statutory and strategic governance. Accountants are required to communicate clearly with clients. They are required to understand and pass on clear information to clients.” (CA ANZ1)
- Read financial documents, follow/ provide instructions to colleagues, communications with clients, identify important information and take action” (CPAA1)

An interesting observation was made by a panelist from ASMIRT. This panelist raised the point that tone of voice to patients was important and language tests did not appear to test nuance of expression. This issue regarding differences in particular professional context training and generic English language proficiency testing will be discussed later in this report.

¹ For practical purposes, the item types are presented here in discrete sections. However, note that PTE Academic uses integrated skills item types. This means that some item types assess more than one skill. For example, Retell Lecture appears in this report under the Speaking section, though it contributes to both the Speaking and Listening scores.

² This new item type was only shown to the second on-line panel

³ This new item type was only shown to the second on-line panel



For Describe Image and Retell Lecture items, panellists focused on situations where relevant information needs to be interpreted and distilled from charts, graphs, technical drawings, client interactions, or meetings and conveyed clearly to a colleague or group. All panellists commented that this ability to verbally summarise and interpret has direct links to tasks in their professions, with the exception of the NAATI panellists. NAATI panellists commented that the task of summarising is not relevant to the act of interpreting, as their role is that of translating with high levels of accuracy more associated with the skills of repeat sentence and read aloud. They did comment however, that the underlying skills required by these item types do have relevance in supporting the ability to interpret.

- “Definitely, presenting data, graphical info is part of daily lives of engineers.” (EA2)
- “Yes. Training staff. Relating information from meetings to colleagues unable to attend the meeting etc.” (AIMS2)
- “Yes – presentation; describing trends, questions; drawings and designs, flows; explaining the benefits of a product; real time simulation results.” (EA1)
- “Yes – it is a major part of a scientist’s job to interpret graphical information.” (AIMS1)
- “Interpreters have an ethical obligation to repeat the whole message. Interpreters are permitted to take notes and generally develop their own style of note-taking. Interpreters would need to synthesise information in a way in order to transfer it to the other language.” (NAATI1)

- “Interpreting data is important to many parts of the various roles pharmacists will perform in Australia. Data sets are usually presented to pharmacists in table or written forms.” (APC1)
- “Understanding patient needs. Effective communication with colleagues and members of the public. So far, this is the most important skill we have seen in these examples.” (APC1)

For Answer Short Question, panellists commented that the task of providing the correct vocabulary word to respond to a request for information was common across professions. Panellists noted, however, that the vocabulary was more likely to be related to the technical knowledge of their profession rather than the general terms assessed in PTE Academic. That said, all panellists expressed a clear need for new members of their profession to have a solid base of vocabulary knowledge in order to communicate clearly within their roles.

- “Yes. Vocabulary of professional terms – very critical. Interview – client knows certain information but want to ask about a tool/material.” (EA1)
- “Not really. Is more of vocabulary of technical terms; however, general vocabulary will be useful for reporting purposes. Yes, it helps to express with the right words.” (EA2)
- “Absolutely – too many to list but needles size, catheter size, air pressure for ventilator, generic names for medications etc.” (NAATI1)

- “I do think it’s a sort of an important skill for accountants, but I think it really sort of largely falls under sort of effective business communication, really just making sure that you know they are sort of able to understand what’s been asked and respond appropriately. I think that’s all just sort of... well, the skills required for effective business communication for companies.” (CPA A1)

The two additional item types of ‘Summarise Group Discussion’ and ‘Responding to a Situation’ were shown and discussed with the on-line panel. These item types are found in other PTE language proficiency tests and are being included as an enhancement to the PTE Academic test.

In the Summarise Group Discussion item, test takers listen to a short discussion among three speakers and then summarise in their own words what was discussed. Each discussion is approximately 1 minute, and test takers are given up to 1 minute 30 seconds to complete the task. Responses are scored on the content of the summary, as well as the fluency and pronunciation of the delivery.

Test takers are expected to summarise the discussion, including the main points and important details, and must understand much of the meaning in order to provide an accurate summary. Listening texts include a variety of unique features, such as fillers and hesitations, false starts, and self-corrections, which add authenticity. They also take place in real-time, meaning the texts are only played once, so test takers must process as they listen as they would do in real-world situations. This task involves understanding the content in order to fully and accurately represent it using coherent and clear speech.

In the Responding to a Situation item, test takers listen to and read a brief scenario and are then asked to respond verbally as if they were in the situation. Test takers have 10 seconds to prepare a response and 1 minute to respond to each situation. Test takers are expected to give relevant, appropriate responses as well as respond clearly and coherently. Test takers are scored on the appropriateness of their response to the given situation, as well as the fluency and pronunciation of their speech.

The panel supported the inclusion of these items to the PTE Academic test:

- “Hearing different perspectives from patients, family and other health-care workers and being able to understand what the issue or preferred resolution is” (APC1)
- “I think you know individuals in the accounting profession would need to be involved in meetings, group discussions and stuff like that, so I think this would definitely be a skill that would be important. And yeah, I think this would be useful to test. I believe I think it does sort of assess the fluency of understanding, bit of the discussions. So I think that will be very helpful.” (CPA A1)
- “Like if your hospital pharmacist is being in situation where you’re hearing, maybe a complaint from the patient and their family member and the nurse and so you know, just a big range of people talking at the same time. You need to actually understand what the issue is and what the preferred resolution is. I can definitely see an application for the types of skills that you’re accessing.” (APC1)

On the importance of the assessed skills of the item type to their professions:

- “Yes, often pharmacists are hearing a lot of discussion at once, especially in interprofessional teams and need to understand the issue and work toward a resolution” (APC1)
- “Yes, I believe this item is important for business cases discussions, meetings, brainstorm, decisions and social interaction.” (CA ANZ1)

When asked if they thought the Summarise Group Discussion item type would be a good addition to the PTE Academic test, the panellists responded positively:

- “Yes. I believe this is an important item not only for business related tasks but also for social interaction” (CA ANZ1)
- “Yes, as this is part of practical language abilities” (CPAA1)

The panellist from APC related the Responding to a Situation item type to their profession:

“I like the unexpected nature of it. So it gets anyone sort of thinking on their feet. And it would be a bit more, you know, conversational. We’ll be able to hear nothing much like what they say, but like you said, they’re not just repeating something. They actually need to understand the situation and provide a response.” (APC1)

Writing

Writing Item types include:

- Summarise Written Text
- Write Essay
- Summarise Spoken Text

For Summarise Written Text, communicating with others via summaries was evident in the panel responses. Multiple panellists highlighted the need to summarise longer written text to simplify the contents for other audiences, such as users, trainees, higher management, or other members of medical care team.

- “Research paper summary to propose new technology/tool. Especially to leadership not in your field. Both in presentation meetings and in email/paper proposals.” (EA1)
- “Rarely, but often it would occur if for example an x-ray report that shows a fractured leg – a summary might simply be “fractured left femur” rather than a detailed imaging report.” (ANMAC1)
- “Yes, these are important skills for translators to transfer written text into English. Although they do not summarise, translators will generally use all the these associated skills to produce a quality translation.” (NAAT1)

In general, the panellists considered that the Summarise Written Text and Summarise Spoken Text item types were more relevant in terms of writing tasks undertaken in their professions:

- “Yeah, there’s such a broad range of what pharmacists can do once they are registered. So they might be doing medication management reviews and they have to look at the whole big long list of things that people are taking and being able to provide different forms of like directions to the patients. Sometimes they even have to write referrals to doctors in a written form. And if they’re a pharmacy owner. They’ll have to understand and, umm, be able to comprehend and write comprehensively for any sort of writing task that would be in the like business realm. So there are quite a lot of applications I can see to pharmacies. (APC1)
- “Summarising issues presented verbally for GP referrals. Hearing different perspectives in a hospital or community practice setting and requiring transcription. Sometimes patients also need written translations” (APC1)

For the Write Essay item, panellists largely commented that traditional argumentative essays were not a requirement of their professions. However, all panellists agreed that the ability to write extended, coherent, logically structured text was necessary to their profession. Panellists were divided on the relevance of spelling, grammar, and punctuation within their field.

For panellists from technical fields, shorthand and technical terminology are valued above precise grammar and punctuation. For NAATI, precise spelling, grammar, and punctuation are fundamental to the role.

- “Very often used. Introducing new technology, research reports, logical way, project management, incident reports.” (EA1)
- “Manuals (users, tech etc), product overviews (product managers), incident reporting, project scope documents (PMO), opportunity management, innovation statements.” (ACS1)
- “On occasion, case notes where logical structure and coherence is important. Incident reporting. Policy writing, position descriptions, but for higher level roles rather than entry level.” (ANMAC1)
- “It is important for Australian based pharmacists to communicate well and directly in a written form” (APC1)

Reading

Reading item types included:

- Gap Fill (Reading)
- Gap Fill (Reading and Writing)
- Multiple Answer Multiple Choice
- Single Answer Multiple Choice
- Reorder Paragraph



Reading item types are entirely closed response. Consequently, the actual task of completing them is unlikely to mirror an exact task in a professional context. The focus of discussion for reading item types was on the underlying skills required to answer each item type, rather than the authenticity of the task itself. Most of the panellists saw some relevancy between these item types and the skills needed their profession. Underlying all reading item types is a requirement for comprehension.

For Gap Fill item types, panellists commented that the skill of using context clues to infer missing or obscure words is necessary for a range of professional roles because real-world written communication can often be flawed or incomplete.

- “Yes. It is critical to understand the context of the document. With increasing use of emails, typo is becoming more common.” (EA2)
- “Interpreting handwriting in patient notes (or on a medication chart which should prompt a r/v).” (ANMAC1)
- The APC panellist did not see relevance in the Gap Fill items in the reading or listening items, as pharmacists must operate with correct information and not infer the meaning of any communications, for example prescriptions.

For Multiple Answer Multiple Choice and Single Choice item types, panellists largely did not find direct links between the item type and professional tasks, but commented that comprehension of meaning, tone, and specific points of information are all important reading skills within their professions.

- “Yes, comprehension of reports and analysis.” (EA1)
- “Yes. Complicated test instructions would need to be read and understood prior to testing.” (AIMS1)
- “Comprehension is very important to the level of English we expect of pharmacists” (APC1)

For Re-order Paragraph, panellists identified a variety of examples of professional tasks that require interpreting disordered narratives or reorganising written information.

- “Translators may sometimes need to reorder information in order for it to make sense in the target language.” (NAATI1)
- “Reorder for coherence – deciphering rough notes post interview with MH/dementia patient to ascertain meaning?” (ANMAC1)
- “Yes, procedures, work flow in the correct sequence is critical in an engineering process.” (EA2)

Listening

Listening item types include a mixture of closed response types, in which test takers select the correct answer, and open integrated item response types, where test takers produce their own response. Listening item types include:

- Summarise Spoken Text
- Multiple Choice Multiple Answer
- Fill in the Blanks
- Highlight Correct Summary
- Multiple Choice Single Answer
- Select Missing Word
- Highlight Incorrect Word
- Write from Dictation

For Summarise Spoken Text, the panels comments were largely in line with their comments related to Retell Lecture. For the majority of the panel, verbal summaries are a common task within the profession, except for NAATI interpreters, who do not carry out this task but access the underlying skills.

- “Yes – Client interactions, generally a summary comment is written into the LIS. Problem solving, incident reporting.” (AIMS1)
- “Yes, all the time. Across most areas, up and down skills level e.g. junior staff to senior management.” (ANMAC1)
- “Agile stand up, virtual meeting (MS Teams), video and podcast briefings, incident management – triage.” (ACS1)

In the TRG panel meetings, more time was given to the productive skills of speaking and writing as they had associated student exemplification at different levels of proficiency for the panels to consider for Activity 2.

The closed nature of items in the reading and listening sections and some restrictions in available time led to less discussion on these two separate skills assessments. As discussed earlier in this paper, many of the items in the PTE Academic test are integrated items and therefore assess receptive and productive skills together. These item types are included in the speaking and writing sections of this report.

Activity 2 Findings – Level of proficiency

Activity 2 was centred on listening to or reading actual test taker responses to the productive skills of speaking and writing where an open-ended response was required. To enhance the continuity of discussion, Activity 2 was carried out for speaking once all the speaking item types had been discussed and similarly at the end of the writing section. A test-taker portfolio method was used, where a number of responses from each test taker were considered. The professional bodies invited to participate in the two TRG panels used the Proficient and Competent visa categories as their language proficiency gatekeepers for migrant entry visa purposes, however they offered useful insights into not only into their expectations of the language proficiency standards in their associated visa categories, but also indicators of language proficiency above and below those standards. The speaking and writing samples were scored items, with known performance outcomes for the test-takers. The scores were not revealed to the panels. We asked them to describe the features of the responses and to consider whether the responses from each test taker would be appropriate evidence of proficiency for their professional body.

Example of rating grid from Panel 1

You will now hear several example responses from the test takers of different level of proficiency. For each one, consider if the level of ability demonstrated in the response would be expected from a candidate who meets the language requirement for your organisation.

Example	Is this level of ability appropriate for your failed?	What qualities of the response stood out for you? Any other comments
1		
2		
3		
4		

Example of rating grid from Panel 2⁴

Item type	Test taker 1 (Superior)	Test taker 2 (Proficient)	Test taker 3 (Competent)
Read Aloud			
Repeat Sentence			
Describe Image			
Retell Lecture			
Summarise Written			
Text			
Write Essay			
Summarise Spoken			
Text			

⁴ As the participants from Panel 2 did not consider the Write Essay to be particularly relevant to their professional body requirements, we excluded examples of essays

All observations were recorded. The key indicators the panels highlighted included:

- Evidence of understanding and engagement with the item prompt, giving reasoned responses
- Ease of dealing with familiar and unfamiliar contexts
- The use of language conventions in terms of grammar and vocabulary accuracy
- Clarity of response. For speaking – accent did not matter – clarity was the key indicator
- Ability to synthesise information
- Fluency of responses

There were no anomalies in their judgements. Therefore, all superior examples were judged as exceeding all other visa category expectations, and this cascaded downwards as the range of exemplar responses were presented to the panels.

Following the TRG event, draft holistic performance descriptors were produced using feedback from the panel in written and oral form. The need for such performance descriptors is apparent. While each professional body is aware of the standards and requirements within their profession, there is a need for clarity and common ground across professions to establish shared interpretations of the DHA proficiency categories in the context of skilled economic migration. As one panellist commented:

The draft performance descriptors were sent to the TRG panel members for review and comment. The performance descriptors shown below are the validated versions at this point.

“The need for such performance descriptors is apparent. While each professional body is aware of the standards and requirements within their profession, there is a need for clarity and common ground across professions.”

Performance descriptors

Superior

The test taker demonstrates very high levels of comprehension, understanding and communicative skills across the four language skills of reading, writing, listening and speaking. Where appropriate language and form can be academic in nature, however the emphasis at this visa category level indicates the ability to absorb textual and spoken information across a wide range of familiar and unfamiliar contexts, demonstrating a very high level and range of comprehension and understanding, necessary language convention structures and vocabulary. The test taker demonstrates almost faultless accuracy when speaking or writing. Their very high level of language proficiency is such that they can understand and carry out specific language instructions promptly and effectively, synthesize a wide range of information and produce highly effective summaries, in both written and spoken forms. They can also use a wide range of linguistic skills to produce fluent reasoned opinions on a given topic in both writing and speaking.

Proficient

The test taker demonstrates good levels of comprehension, understanding and communicative skills across the four language skills of reading, writing, listening and speaking. The emphasis at this visa category level indicates the ability to understand textual and spoken information taken from familiar and unfamiliar contexts, demonstrating a good level and range of comprehension and understanding, and

utilizing usual language convention structures and vocabulary. The test taker demonstrates good levels of accuracy when speaking or writing, making a few errors but which can be understood with ease. Their level of language proficiency is such that they can understand and carry out language instructions promptly and effectively, synthesize information and produce effective summaries, in both written and spoken forms. They can also use a range of linguistic skills to produce fluent reasoned opinions on a given topic in both writing and speaking.

Competent

The test taker demonstrates adequate levels of comprehension, understanding and communicative skills across the four language skills of reading, writing, listening and speaking. The emphasis at this visa category level indicates the ability to understand the main meaning of textual and spoken information taken from familiar and unfamiliar contexts, demonstrating general comprehension and understanding, and utilizing some language convention structures using a limited range of vocabulary. The test taker can be understood when speaking or writing, however they do display technical inaccuracies, some hesitancy and lack of fluency. Their level of language proficiency is such that they can in general understand and carry out language instructions effectively, comprehend the main meaning of information and produce moderate summaries, in both written and spoken forms. They can use a limited range of linguistic skills to produce opinions on a given topic in both writing and speaking.

Vocational

The test taker demonstrates working levels of comprehension, understanding and communicative skills across the four language skills of reading, writing, listening and speaking. The emphasis at this visa category level indicates the ability to engage with of textual and spoken information taken from familiar contexts, demonstrating some comprehension and understanding, and utilizing a few language convention structures using a very limited range of high frequency vocabulary. The test taker can be mostly understood when speaking or writing, however they do display many technical inaccuracies, hesitancy and lack of fluency. Their level of language proficiency is such that they can understand and carry out some language instructions, comprehend some meaning of information and attempt summaries, in both written and spoken forms. They can use a very limited range of linguistic skills to produce opinions on a given topic in both writing and speaking.

Discussion

The research questions that this study set out to investigate were:

- Are PTE Academic test items relevant and valid for assessment of English language proficiency in the domain of professional body visa entry requirements into Australia?
- Are the skills elicited by the PTE Academic item types relevant and valid within professional bodies in Australia?
- Do PTE Academic performance standards as categorized by Australian visa categories match professional bodies expectations of English Language proficiency?

- Can Proficiency Level Descriptors be developed to identify the key competencies elicited in English language proficiency tests at each Australian visa classification?

The TRG panels were comfortable that the construct of the PTE Academic is based on the Common European Framework of Reference (CEFR) and includes the assessment of speaking, writing, reading and listening skills, in order to be able to comprehend and understand English language in a variety of contexts and be able to effectively communicate through speaking and writing. They did not offer any alternative basis for a professional domain language test compared to any other domain.

Most panel members were also comfortable with the purpose of a language proficiency test being agnostic to a particular profession. This study involved the participation of twelve representatives from nine professional bodies. Some of these professional bodies may have credentialing programmes including the assessments of their own particular Knowledge, Skills and Judgements (KSJs), as well as the assessment of performance in practice. It was clear from discussion that the panels did not expect general language proficiency tests to be so focused on their particular professional context. Indeed, they were supportive that a test such as **PTE Academic made test-takers engage with familiar and unfamiliar contexts and vocabulary outside of a particular professional setting.** Links made to professional job roles and tasks were welcomed, however the panels were also looking for contexts and tasks that were relatable in a general sense for people to engage in Australian society.

The representative from ASMIRT raised the issue that in their medical roles, radiographers would need to demonstrate empathy and nuance when talking to patients. This is a good example of the essential difference between a general language proficiency test and a KSJ test. Such differences were acknowledged by the panels.

The first two research questions focused on the relevance and validity of the PTE Academic item types to the professional bodies in attendance and also whether skills required in any of their professional body job roles and tasks could be elicited through performance of the item types.

There was considerable support from the TRG panels of such relevance and validity. The findings section of this report provides multiple examples of responses from the panel. These examples are a sample of the comments recorded from the meetings.

The confirmation of relevance was welcomed, however not entirely surprising. Although the PTE Academic test is used in professional and study domains, the test-taking cohort are at an under-graduate or post graduate level and so expectations are that tasks are realistic in terms of demonstrating skills found in study, office, business or public/patient environments. There is also the need to discriminate between levels of performance, and so there is a requirement to include a range of linguistic challenges and levels of difficulties and demand.

As the purposes of language proficiency tests have changed over time, so has the contexts of item types. The only productive item type that had only partial support in the TRG panel meetings was the write

essay. Some of the professional bodies felt that extended writing in this form was not a feature of many of their professional roles. They did however acknowledge the importance of developing sound underpinning writing skills either in short summarising or longer forms.

The second panel were presented with two additional item types soon to be included in the PTE Academic test. These two items, 'Summarise Group Discussion' and 'Respond to a Situation' are spontaneous speaking items and are purposed to be dynamic and complex in nature, requiring high levels of comprehension, understanding, synthesis and oral skills. The TRG panel was highly supportive of the inclusion of these two item types. They considered them to be very effective in terms of the language skills they assess, and also their relevance to professional contexts.

The PTE Academic test assesses receptive and productive language skills through a combination of single and integrated skills. Whereas the items included in the assessment of productive speaking and writing skills, via single or integrated skills were in the main considered to be highly relevant to the professional bodies, the single skills items assessed in reading and listening, in the form of closed response item types received mixed views. The TRGs found it difficult to assimilate some of the item types and response mechanisms to professional tasks, however they acknowledged the importance of the measurement of reading and listening comprehension skills. Item types such as re-ordering paragraphs and gap fill items were seen as relevant to a number of professional bodies, multiple choice items far less so.

In terms of assessment, there are going to be trade-offs in terms of the authenticity of item types and measurement qualities. Although open-ended items have high levels of relevance and authenticity, they cannot be relied on to assess complete language constructs. That is why high stakes assessments use such item types.

High stakes assessments have reliability, validity and fairness responsibilities. They also have to operate in the context of unavoidable real-world constraints. Although evidence of validity is essential, regulatory assessment procedures are never designed to only consider validity.

Instead, the objective underlying assessment design is to optimise validity; typically, to accommodate a broad profile of intended purposes, and to recognise a wide range of operational constraints (Newton, 2017b).

Therefore, assessment design is fundamentally concerned with trade-off and compromise. These include:

- the desire to increase the reliability of results by increasing the number of assessment tasks, versus the ability of candidates to sustain concentration and effort when the duration of an assessment event is too long;
- the desire to measure complex skills authentically using tasks that mirror real-world situations, versus the ability of assessors to evaluate complex performances with consistent accuracy; and
- the desire to measure all of the elements identified within a proficiency specification, versus the ability of an assessment community even to reach consensus over criteria for judging certain elements. (Ofqual, 2017)

In terms of the first two research questions of this study, the relevance and validity of the PTE Academic test for the purpose of testing English Language proficiency in professional domains was endorsed and supported by the TRG panels providing feedback and evidence in this report. The PTE Academic test was considered fit for purpose in terms of testing a broad construct of essential language skills with high levels of accuracy and reliability and also including highly relevant and valid item types pertinent to the assessment of the participating professional bodies and reflective of many processes, procedures and tasks evident in their professional job roles.

The second two research questions were concerned with DHA visa performance standards.

The TRG panels endorsed the DHA visa categories performance standards in relation to a sample of PTE – Academic test takers responses that had met the test scores necessary for a range of visa classifications.



The link that was made in this research study was to verify that the expectations of English language competency that professional bodies expected was achieved through actual test-taker performances in the PTE Academic test. Apart from this verification, there were two other significant outcomes and findings in this phase of the research.

The first outcome was the development of Visa Proficiency Level Descriptors. These do not currently exist in the context of DHA visa categories, and therefore there is no public information regarding what visa classifications, in assessment outcomes, use look like in practice, particularly in the context to professional and occupational usage of English language. The emergent Performance Descriptors were aligned to the PTE Academic test, however, as holistic statements they provide a working description of performance standards that should be applicable across test providers.

The Performance Descriptors presented in this report are the outcomes of oral and written discussion, comments and feedback given by nine professional bodies in Australia, including a number of significant economic and societal areas such as Accountancy, Health and Medical services and Technology. As such, they add to a professional domain validity argument.

The second outcome was the desire for professional bodies to be included in the evaluation of English Language proficiency tests and their associated standards. As key stakeholders, the TRG panel members were willing and keen participants and contributors to this research study.

It is clear that if the differing purposes of English Language Proficiency tests are to monitored, evaluated and enhanced, then a wide range of constituent stakeholders should be involved, including policy makers, testing organisations, universities and professional bodies.

Conclusion

The findings of the TRG activities and follow-up survey indicate the PTE Academic is relevant to the tasks and underlying skills required by a range of professional contexts in Australia. Though PTE Academic was developed to assess English language skills required by higher educational settings, the findings of this report indicate that many of the tasks have direct links to real-world tasks undertaken by new members of professional fields, and that these tasks tap into important underlying skills valued across a range of professions. By engaging directly with professional bodies, this study provides rich exemplification of the practical links between the language proficiency assessed by PTE Academic and required by professional organisations in Australia. This has enabled us to develop descriptors of DHA language proficiency categories that are rooted in the needs and requirements of professional bodies.

This form of validation research study is not intended to be one-off activity. The PTE Assessment Research and Validity team will be undertaking further research programmes to support the ongoing evidence base on the domain relevance of PTE Academic test in both academic and professional contexts.

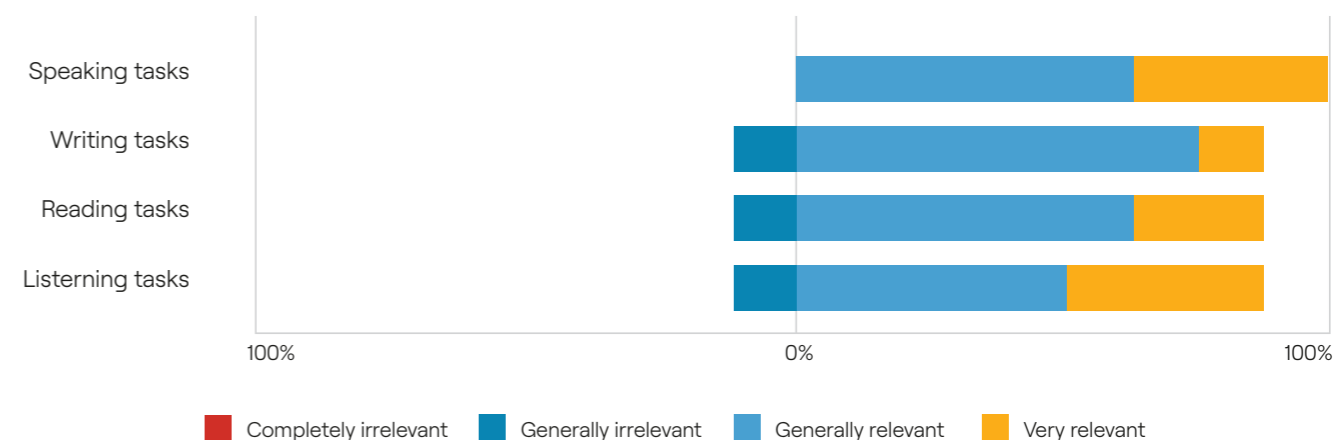
Survey responses

A follow-up survey was sent to participants to capture their background and overarching opinions on PTE Academic. Of the 8 panel participants in the first TRG, 6 responded to the survey. Of the 4 participants in the second TRG, 2 responded to the survey.

The questions on PTE Academic test relevance are shown below.



14. How relevant to your profession are the tasks in the PTE Academic test?



The respondent who responded that the Writing, Reading, and Listening tasks were “generally irrelevant” was a panellist from NAATI. As discussed in the Activity 1 findings, interpreters and translators carry out very different professional tasks to

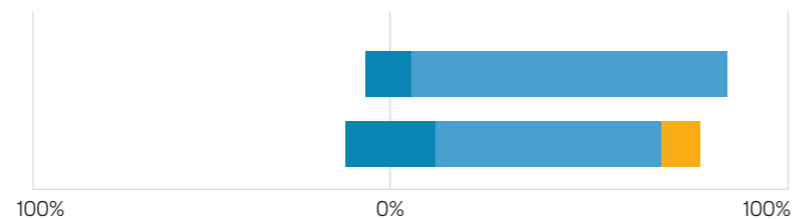
other professions, and PTE Academic is not designed as a test of interpretation and translation. However, the same panellist agreed in the following question that PTE Academic assesses important underlying skills for their profession.

7 out of 8 respondents agreed the PTE Academic item types are relevant to their profession.

15. To what extent do you agree with each statement below?

PTE Academic assesses language skills that are important for a newly qualified member my...

PTE Academic scores are a useful indicator of language proficiency for my profession.



■ Strongly disagree
 ■ Disagree
 ■ Neutral
 ■ Agree
 ■ Strongly agree

Of the 8 respondents, 7 of them agreed the PTE Academic assesses important skills for their profession and is a useful indicator of proficiency for their profession. Of the 8 respondents, 6 of them agreed PTE Academic scores are a useful indicator of language proficiency for their profession. One respondent remained neutral for both questions, although offering supportive opinions and exemplification throughout Activity 1 and 2. The same applied also to the second question.

Acknowledgements

We would like to thank the following professional organisations who participated in this research. No organisations or individuals were paid for their involvement.

- Australian Nursing & Midwifery Accreditation Council
- National Accreditation Authority for Translators and Interpreters Ltd
- Engineers Australia
- Australian Computer Society
- Australian Institute of Medical and Clinical Scientists
- Certified Public Accountants
- Chartered Accountants
- The Australian Society of Medical Imaging and Radiation Therapy
- Australian Pharmacy Council



References

AERA, APA, and NCME (2014). Standards for Educational and Psychological Testing (6th edition). Washington, DC: American Educational Research Association.

Council of Europe. (2009). Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching assessment (CEFR). A manual. Language Policy Division.

Egan, K., Scheider, M., & Ferrara, S. (2012). Performance Level Descriptors: History, Practice, and a Proposed Framework. In G. J. Cizek (Ed.), Setting performance standards: Foundations, methods, and innovations. Routledge.

Frost, K., Fan, J., Blain, H., & McNamara, T. (2018). Stakeholder perceptions of the use of PTE Academic for the purpose of skilled migrant selection in the accounting profession of Australia. Language Testing Research Centre, University of Melbourne.

Hambleton, K. (2012). Setting Performance Standards on Educational Assessments and Criteria for Evaluating the Process. In G. Cizek (Ed.), Setting performance standards: Foundations, methods, and innovations. Routledge.

Kane, M.T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50 (1), 1–73.

Knoch, U., & Pitman, A. (2019). Examining the relevance of the Pearson PTE Academic test tasks for professional nursing registration – a stakeholder perceptions study. Language Testing Research Centre, University of Melbourne.

Macqueen, S., Pill, J., & Knoch, U. (2021). Trust the test: Score-user perspectives on the roles of language tests in professional registration and skilled migration. *Papers in Language Testing and Assessment*, 10(1), 49–69.

Newton, P.E. (2007) Assessment-A system fit for purpose? *Westminster Education Forum Keynote Seminar*

Newton, P.E. (2017) An Approach to understanding validation arguments. Ofqual.

Shaw, S.D. and Crisp, V. (2012). An approach to validation: Developing and applying an approach for the validation of general qualifications. *Research Matters: A Cambridge Assessment Publication*, Special Issue 3, 1–44.

Shaw, S. and Crisp, V. (2015). Reflections on a framework for validation – Five years on. *Research Matters: A Cambridge Assessment Publication*, 19, 31–37.

Appendix A – Worksheet Exemplar

Read Aloud

Name _____

Organisation _____

This item type asks test takers to read written text aloud.

Is this item type similar to any tasks in your profession field?

This item type is designed to assess ability to read the content of a passage accurately, and to communicate the content clearly.

Are those skills used in your professional field? Does communication in your field use these skills?

You will now hear several example responses from the takers of different level of proficiency. For each one, consider if the level of ability demonstrated in the response would be expected from a candidate who meets the language requirement for your organisation.

Example	Is this level of ability appropriate for your failed?	What qualities of the response stood out for you? Any other comments
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		



About PTE

Learn about who we are and what we do.

PTE is a world-leading provider of secure English language tests. Pearson Test of English (PTE) provides secure English language testing for study applications worldwide, and for visa applications for work and migration in Australia, the UK, New Zealand, and Canada. Every year, tens of thousands of people trust PTE to help them rapidly prove their English proficiency – and to open doors to their future lives and careers.

