Preparing for the NAATI examination: options and issues for English proficiency screening

Final Report to the National Accreditation Authority for Interpreters and Translators (NAATI)

Catherine Elder, Ute Knoch, Hyejeong Kim
Language Testing Research Centre
The University of Melbourne
February 2016
Contents

0. Executive summary ........................................................................................................................................... 4

1. Background ....................................................................................................................................................... 5

   What do we mean by “basic requirements for adequate bilingual work?” ..................................................... 6

   The CEFR ........................................................................................................................................................... 7

2. Statement of requirements for a language screening test ................................................................................ 9

   Validity ............................................................................................................................................................... 9

   Reliability ......................................................................................................................................................... 10

   Comparability ................................................................................................................................................ 11

   Utility ............................................................................................................................................................... 11

   Equity ............................................................................................................................................................. 12

3. Methodology .................................................................................................................................................... 14

4. Findings ............................................................................................................................................................. 14

5. Discussion ......................................................................................................................................................... 14

   Validity ............................................................................................................................................................... 14

   Reliability ......................................................................................................................................................... 15

   Comparability ................................................................................................................................................ 15

   Utility ............................................................................................................................................................... 16

   Time .................................................................................................................................................................. 16

   Delivery ........................................................................................................................................................... 16

   Pricing ............................................................................................................................................................. 16

   Quality of feedback ......................................................................................................................................... 17

   Equity ............................................................................................................................................................ 17

6. Conclusion and recommendations .................................................................................................................. 18

7. References ........................................................................................................................................................ 21

8. Appendices ......................................................................................................................................................... 24

   Appendix A. Entry requirements for accredited translator and interpreter courses in Australia ............... 24

   Appendix B. Global scale descriptors for Common European Framework of Reference (CEFR) levels .......... 29

   Appendix C. Findings of shortlisted tests ....................................................................................................... 30

      1. Academic English Screening Test/Post-entry Assessment of Academic Language ............................... 30

      2. Aptis ........................................................................................................................................................ 32

      3. CaMLA English Placement Test (computer-based) ............................................................................... 35

      4. Diagnostic English Language Needs Assessment (DELNA) Screening ............................................... 37

      5. DIALANG ............................................................................................................................................. 38

      6. Education First Standard English Test (Plus) ....................................................................................... 40

      7. Oxford Online Placement Test ............................................................................................................ 42

      8. Password ............................................................................................................................................. 44
9. Progress .................................................................................................................................................... 46
10. Versant English Placement Test.............................................................................................................. 48
11. Versant English - Speaking ...................................................................................................................... 50
12. Versant English - Writing ........................................................................................................................ 52

Appendix D. Samples of score reports of shortlisted tests............................................................................... 56

1. Academic English Screening Test/Post-entry Assessment of Academic Language ............................... 56
2. Aptis .......................................................................................................................................................... 56
3. CaMLA English Placement Test (for institutions) ...................................................................................... 57
4. Diagnostic English Language Needs Assessment (DELNA) Screening ....................................................... 58
5. DIALANG (e.g., Listening) .......................................................................................................................... 58
6. Education First Standard English Test (Plus) ............................................................................................. 59
7. Oxford Online Placement Test (for Institutions) ....................................................................................... 59
8. Password ................................................................................................................................................... 60
9. Progress .................................................................................................................................................... 60
10. Versant English Placement Test .............................................................................................................. 61
11. Versant English-Speaking ........................................................................................................................ 62
12. Versant Writing ....................................................................................................................................... 64
0. Executive summary

In late 2015 the Language Testing Research Centre at the University of Melbourne was commissioned to undertake a feasibility study exploring the possibilities for preliminary language screening of those applying to sit for the certifying examinations run by the National Accreditation Authority for Interpreters and Translators (NAATI). The study was triggered by a recommendations in a report prepared by Hale, Garcia, Hlavac, Kim, Lai, Turner & Slatyer (2012) which proposed a radical reform of the current NAATI examination system. As part of the exam reform it was recommended that an online screening test in English and other relevant languages be made available to those seeking professional certification as a means of ascertaining that they had achieved the level of bilingual competence needed to engage in the language transfer activities germane to the translator and interpreter (hereafter T&I) profession. The study was conducted in two parts, the first of which focused on English proficiency screening. The second was concerned with screening proficiency in the range of additional languages serviced by NAATI, as well exploring additional possibilities for testing in languages other than English, to be potentially managed by NAATI for purposes other than pre-certification screening.

The current report focuses on the first part of the study exploring options and issues surrounding English proficiency screening. It begins by providing some background for the project, including consideration of a) the basic language proficiency requirements for bilingual work as outlined in recent reports commissioned by NAATI, b) the English language levels required of those applying for professional accreditation in Australia based on an overseas qualification and c) the published English proficiency entry requirements for those undertaking T&I training in accredited NAATI courses. It highlights the importance of professional consensus about not only what kind of English proficiency is needed but also about how much proficiency is enough for effective T&I work. It also considers the relevance of the widely used Common European Framework of Reference (CEFR) to determinations of levels of language proficiency and to making comparisons between different testing systems.

The report then outlines a set of requirements or principles that can be used as a basis for evaluating available English language tests for the intended pre-accreditation screening purpose. The test requirements of Validity, Reliability, Comparability, Utility and Equity are explained and issues of particular relevance to the current project are highlighted.

The methodology adopted by the researchers is then described and findings are presented in the form of detailed descriptions of 12 shortlisted online English tests following the parameters identified in the Statement of Requirements and making reference to relevant technical reports and research articles. The tests in questions are the Academic English Screening Test developed at the University of Melbourne, the Aptis Test developed by the British Council, the CaMLA English Placement Test developed by the Cambridge/Michigan group, the Diagnostic English Language Needs Assessment (DELNA) Screening tool developed at the University of Auckland, The Dialang test developed at the University of Lancaster, the Education First Standard English Test developed by Education First (an international education company), the Oxford Online Placement Test developed by Oxford English Testing, the Password test developed at the Centre for Research in English Language Learning and Assessment at the University of Bedfordshire and four tests developed by the Pearson company: the Progress test, The Versant English Placement Test, Versant English (Speaking) and Versant Writing. A discussion section considers the strengths and weaknesses of the various tests and their fitness (or otherwise) for the intended purpose.

The report concludes by identifying three possible options for pre-accreditation English proficiency screening, and makes the following recommendations to NAATI:
1. That NAATI consult with T&I experts to consider available language proficiency frameworks (such as CEFR) and language testing regimes (such as IELTS) and determine what level or levels they deem most suitable as an English proficiency benchmark for the T&I profession;

2. That NAATI and T&I experts consider the following test options for English proficiency screening as discussed above;
   
   Option A: Versant English Speaking for all applicants for the interpreting exams and Versant Writing for all applicants for translating.
   
   Option B: AEST (or DELNA Screening) for all NAATI applicants followed by a recommendation to sit Versant English Speaking (for interpreters) or Versant Writing (for translators) for those who score above a given threshold.
   
   Option C: A customised version of AEST which includes a listening component to be used with or without the Versant English Speaking and Writing add-ons.

3. That NAATI negotiate with the relevant agency/ies to run a year-long trial of the chosen option with both T&I students and applicants for the NAATI exam for whom English is a second language;

4. That NAATI conduct a standard-setting exercise with members of the T&I profession to establish recommended minimum score thresholds on the chosen test or tests above which candidates might be considered ready to sit the relevant exam;

5. That NAATI develop a language proficiency policy (in relation to English and other languages) which sets out language proficiency expectations for entry and exit from accredited T&I courses and explains the rationale for preliminary language proficiency screening as a means of ascertaining readiness to take the NAATI exams, as well as the nature of the chosen test/s and how scores should be interpreted and acted on;

6. That students register voluntarily with NAATI to take the chosen test or tests for a fee (consistent with the costs of NAATI's site-licensing arrangements with the relevant agency);

7. That immediate feedback on test performance be provided to candidates where possible and used by NAATI for research purposes only;

8. That the English screening program, once implemented, be reviewed on a regular basis to enable adjustments to policy and practice in light of new developments in the language testing field.

1. Background
The current project elaborates on one of the recommendations contained in the report, ‘Improvements to NAATI testing: development of a conceptual overview for a new model of NAATI standards, testing and assessment’, prepared by Hale et al. (2012) for the National Accreditation Authority for Interpreters and Translators (NAATI) in November 2012. It specifically addresses the third recommendation as set out on page 7 of the report, namely: “That NAATI select (or devise) an on-line self-correcting English proficiency test to be taken by potential candidates for a fee, as part of the non-compulsory preparedness stage, as outlined in sections 2.3 and 3.1. 4.”
In Section 2.3 of the report a five-stage accreditation process for interpreters and translators is proposed with the preliminary objective of stage 0 being to aid potential candidates in understanding “the basic requirements for adequate bilingualism before they invest more time and money into investing any further in the process.” (p. 39). As part of this consciousness-raising process the report proposes that “an online test of English proficiency be developed or made available that can be self-administered (for a fee) and autocorrected upon completion.”(p.40) It is also advocated that applicants who fail to attain a specified minimum score on this test (or recognized equivalent level of performance on other recognized tests or formal English study program) be advised of avenues for further English development before reattempting the test after a twelve-month period.” (p. 40)

In this report we consider what might be considered the basic requirements for establishing whether a candidate has adequate English for bilingual work in light of accepted standards for valid, useful and equitable assessment in the language testing field. We then review a number of tests that meet, or come close to meeting, these requirements which include feasibility issues associated with test delivery and cost. Finally we lay out some options for NAATI to consider as well as a set of recommendations for further action and research that might be undertaken in relation to the project.

What do we mean by “basic requirements for adequate bilingual work?”

An assumption underlying the project is that a common understanding can be reached within the translator-interpreter (hereafter T&I) profession in Australia about not only the components of English proficiency needed for interpreting and translating work but also the level of proficiency required. In other words, some agreement is needed about not only what knowledge and abilities are needed but also about how much language proficiency is enough. Brief consideration will be given to both of these questions below.

Consideration has been given to the kind of English required for interpreters and translators working in an English-medium environment in two recent reports commissioned by NAATI as part of Phase 2 of the Improvements to NAATI testing (INT) Project. The first, ‘NAATI Translator Certification: Knowledge, Skills and Attributes’ (https://www.naati.com.au/PDF/INT/Translator_KSA_Paper.pdf) is available on the NAATI website and the second NAATI Interpreter Certification: Knowledge, Skills and Attributes has been provided to the authors of this report in draft form. Both refer to ‘language proficiency enabling meaning transfer’ as distinct from language proficiency for communication, which may not be sufficient for meaning transfer across languages. Such language proficiency is described in both documents as building on vocabulary knowledge, grammar knowledge and idiomatic knowledge, or more specifically:

- Register and style appropriate to audience or end use,
- idiomatic collocation, lexis and syntax,
- vocabulary, including non-standard forms and foreign loan words
- grammatical structures, including tense, mood, aspect, voice, gender and grammatical numbers
- plain language (mentioned for interpreters only)
- punctuation and paragraphing (mentioned for translators only)
- Orthography and its variations
- Word and language usage appropriate to audience and context
- Pragmatics, i.e. the underlying meaning of language in context
- textual devices that create cohesion and coherence [mentioned for translators only],
- regional varieties, including differences in dialects and language varieties(e.g. French spoken in France, Canada and other countries) [mentioned for interpreters only]
The report also refers to ‘language trends knowledge’, meaning awareness of changes in language usage in different communities and over time and of diversity of text types and presentation modes associated with new media and technology. This awareness could be classified as metalinguistic knowledge, rather than language proficiency per se, but the presence or absence of such knowledge might well be apparent in the way a written text or spoken text is understood and rendered in another language.

Other competencies detailed in the two reports, such as Intercultural Competence and Transfer Competency, are beyond the scope of language proficiency assessment for the purposes of this project.

While the range of knowledge and skills listed in the two reports suggests the need for a high level of general proficiency in both languages before a candidate would be ready to sit a NAATI exam, there is no general agreement about the precise level of proficiency needed in English (or in LOTE, which will be the subject of a subsequent report by the authors). NAATI currently requires a bandscore of IELTS 7.0 from applicants applying for professional accreditation based on an overseas qualification and IELTS 7.5 for those applying for the levels above Professional. While no standard is set for Paraprofessional level, because accreditations at this level are not awarded based on an overseas qualification, there has been a linkage through the NAATI-approved course pathway of Paraprofessional level with IELTS 6.5 as a recommended entry requirement for students. The basis for these associations is unclear. Moreover, a brief review of published entry requirements for accredited T&I courses in Australia conducted by the researchers shows that in practice entrance standards vary widely and are as low as IELTS 5.5 in some cases (see Appendix A for further information). There is also variation in requirements for test sub-components, with some institutions requiring a minimum bandscore for particular sub-skills as well as an overall average. More importantly, for the current project, there is no explicit statement on course exit standards which would arguably be the appropriate benchmark for determining linguistic readiness to sit the NAATI exam.

Agreement on how much proficiency is enough will be necessary for the purposes of establishing a threshold on a given test or tests at or above which a candidate could be considered linguistically proficient enough to sit the NAATI exam. Mechanisms for reaching such agreement will be discussed later in this report. Another issue that will need to be considered by the profession is whether such thresholds should be different for translating and interpreting given that the former emphasizes reading and writing, whereas the latter is more focussed on listening and speaking. A further question is whether different thresholds need to be set for different levels of the exam.

**The CEFR**

In making such determinations it may be useful to consider the Common European Framework of Reference (CEFR), a set of guidelines for setting learning objectives, developing teaching materials, and making comparisons between systems. The CEFR summarises the scope of a global consensus on levels of functional communicative language ability, expressing these in terms that are meaningful to users. The CEFR scale makes distinctions between three broad learner levels (A (basic), B (independent) and C (proficient), each of which is subdivided into two finer gradations (A1 and A2, B1 and B2 etc.). Since it was first published (North, 2000) it has become widely influential amongst policy makers in providing a reference point for comparing language qualifications designed for different purposes. The level most pertinent in terms of setting minimum standards for
the T&I profession is perhaps C1, the lower level of the Proficient User scale, expressed in global terms as follows:

**Global** At the C1 level of proficiency, students should be able to understand a wide range of demanding, longer texts, and recognise implicit meaning. They can express themselves fluently and spontaneously without much obvious searching for expressions. They can use language flexibly and effectively for social, academic and professional purposes. They can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.

A full set of global level descriptors are provided in Appendix B. More detailed descriptors are also provided for each different skill, such as listening or reading, and for each area of knowledge, such as vocabulary and grammar.

It should be noted that the CEFR is not a test instrument but rather a set of criterion level statements which can be used by test designers to help users locate a test in terms of its likely appropriateness for learners at particular ability levels and also to explain the meaning of scores on a number of different measurement scales in terms that will be meaningful for score users. Given the CEFR’s currency amongst policy makers as well as language teachers and learners, test developers are at pains to formally align their tests to the CEFR level, using variety of empirical procedures, some of which are outlined in a standard-setting manual (https://www.coe.int/t/dg4/linguistic/Source/ManualRevision-proofread-FINAL_en.pdf). For example, the relationship between the CEFR and the International English Language Testing Service (IELTS) test, which is the most widely used English proficiency admissions test used in Australia, are represented in the following chart.

*Figure 1: A CEFR-IELTS equivalence chart*  
(Source: www.ielts.org/researchers/common_european_framework.aspx)
It can be seen from this chart that the IELTS test levels have been equated to the Levels A2 to C2 on the CEFR scale, with an IELTS score of 5 - 6.5 falling at various points along the B2 level ability spectrum, higher scores (7-8) falling along the C1 ability band and the highest score (9) in the C2 range. The IELTS scoring scale is thus finer in its gradations than the CEFR levels. When considering the appropriate threshold level/s on any English test selected by NAATI for self-screening purposes, it may therefore be more precise to talk in terms of its relationship to particular IELTS scores rather than to its corresponding level on the CEFR. This issue of equivalence will be considered further below. An alternative framework worthy of consideration would be Pearson’s recently developed Global Scale of English http://www.english.com/gse#.VrwdaREz7ww, which is broadly linked to the CEFR but describes language levels in more granular terms and lays out precise learning objectives for specific domains such as Professional English.

2. Statement of requirements for a language screening test

The Hale et al. (2012) report outlines a set of evaluation principles proposed by Clifford (2001) which are considered relevant to interpreter-translator assessment. Although the principles for evaluating the soundness of a particular test or test use have in recent years been framed somewhat differently and/or with different emphases by scholars working specifically in the field of language testing (e.g., see Bachman & Palmer, 2010; Chapelle, 2012), the available frameworks generally cover similar considerations. For the sake of consistency with Hale’s report we have applied Clifford’s principles to our own test scoping work, but with some modifications. The meaning of each of Clifford’s principles as they pertain to the proposed the English language screening project have been slightly reworded and are explicated below.

Validity

An instrument can be seen as valid if it measures what it was designed to measure, that is, if it yields scores that allow test users to make accurate inferences about the targeted ability.

In the case of this project the targeted ability is proficiency in English and it is important that the selected testing instrument (or instruments) contains tasks or items that elicit the appropriate range of English skills, whether written or spoken, needed as a basis for undertaking work or further training in interpreting or translating. The language knowledge and skills component of the two NAATI reports are an important point of reference in determining the validity of the test in question. Establishing whether a test is a good measure of the targeted language abilities is partly a matter of professional judgment but should be buttressed by theoretical rationales outlining the nature of the test construct (in our case a particular view or model of language ability) and the thinking behind the design of test items. It is desirable that designers of the instrument also furnish empirical evidence that the scores yielded by the test correlate well with those of other English proficiency measures deemed to be measuring relevant skills and/or, better still, that the test scores accurately predict English language competence demonstrated in the target language use domain.

To be valid for its intended purpose it is of course critical that the relevant English test/s be pitched at the appropriate level of difficulty. In particular, evidence will be needed to ascertain that the minimum score requirement on any test/s deemed fit for the specified purpose does indeed correspond to what the T&I profession considers an appropriate baseline for embarking on the credentialing process (the “how much proficiency is enough” issue flagged earlier in this report).

Since, according to the conceptual model presented by Hale et al. (2012), the screening test is intended as the first step in the credentialing process to be followed in most cases by further language-specific training, the threshold level on the relevant English test needs to be high enough to ensure that candidate on or at that level are sufficiently competent to benefit from this training. It
will be necessary to achieve professional consensus about the minimum proficiency requirements for each level and for each language skill. Whether the minimum requirements should differ for interpreters and translators or be the same for all applicants is also a matter for professional determination.

It may be desirable to set more than one standard depending on the level of NAATI exam for which the student is applying. The mechanism for carrying out such a standard-setting exercise (Cizek & Bunch, 2007) will be discussed later in the report but the point to be emphasized here is that basing the set standard on a systematic and defensible process is a key validity requirement.

Reliability

An instrument is reliable if its items and scoring procedures function consistently and it provides stable results from one administration to another in comparable conditions of use.

Reliability has been variously described as a necessary condition for validity or as part and parcel of validity (Davies et al. 1999, Chapelle, 2012). While all tests contain some measurement error, it is important that any such error or randomness in test results is reduced to a minimum. Statistical evidence should be available to support any test reliability claim (AERA/APA/ACME 1999).

In choosing or developing an appropriate English test it will be important to pay attention to the internal consistency of the test items to ensure that they are all working together to measure the ability in question. Various statistical procedures are used for this purpose and reliability is usually reported as index on a 0-1 scale with 1 indicating a perfectly reliable test. Since all tests contain measurement error, levels above 0.7 are regarded as acceptable for low stakes tests and above 0.8 or 0.9 for high stakes tests (Fulcher & Davidson, 2007). Scoring procedures for open-ended responses will also need to be consistent, with evidence of intra- and inter-rater/assessor agreement provided where relevant. Many tests designers also report the standard error of measurement (SEM) associated with test results and that is, the average margin of error around the reported score on the measurement scale which should be taken into account when interpreting the test result. Since all tests contain measurement error, levels above 0.7 are regarded as acceptable for low stakes tests and above 0.8 or 0.9 for high stakes tests (Fulcher & Davidson, 2007). Scoring procedures for open-ended responses will also need to be consistent, with evidence of intra- and inter-rater/assessor agreement provided where relevant. Many tests designers also report the standard error of measurement (SEM) associated with test results and that is, the average margin of error around the reported score on the measurement scale which should be taken into account when interpreting the test result. Since all tests contain measurement error, levels above 0.7 are regarded as acceptable for low stakes tests and above 0.8 or 0.9 for high stakes tests (Fulcher & Davidson, 2007). Scoring procedures for open-ended responses will also need to be consistent, with evidence of intra- and inter-rater/assessor agreement provided where relevant. Many tests designers also report the standard error of measurement (SEM) associated with test results and that is, the average margin of error around the reported score on the measurement scale which should be taken into account when interpreting the test result.

If a test is unreliable, it means that factors other than the candidate’s ability in the target language may be interfering with test scores. The test method is a common source of measurement error: for example if the majority of test-takers can arrive at an answer through guessing rather than by drawing on the ability of interest, then the test items will not function consistently or yield accurate information. Lack of standardization in test conditions is a further potential source of error. Since the test is to be delivered online, delivery mechanisms should be organized to ensure consistency across candidates. For example it is important that the candidate cannot manipulate test conditions, such as timing, in a way that might interfere with the consistency of test results, unless the time taken by the test candidate is itself factored into the final estimate of language ability.

Evidence should also be available to ascertain that any alternate forms of the test, should these be available, yield comparable information about candidates. The availability of parallel forms is itself a desirable quality given that the screening test may be taken more than once by a candidate and there is a risk of a practice effect artificially boosting test scores if the same tasks or items are attempted more than once. This is not however a relevant consideration for computer-adaptive tests which draw on a large bank of pre-tested items and are designed in such a way a repeater is unlikely to encounter the same set or sequence of items in two successive rounds of testing.

---

1 For example, if a score falls within the SEM margin on either side of the ‘Pass’ or ‘Competent’ threshold on a test it may need to be reviewed.
Comparability
An assessment can only be deemed comparable to another measure if it can be
demonstrated to be measuring similar abilities and if the level on the assessment has been
formally linked to a corresponding level on the other measure.

Our definition of comparability is quite different from the use of this term by Clifford (2001) as cited
in the Hale et al. report, where it covers issues of assessor training and consistency of scoring that
are normally subsumed under reliability. The issue of comparability that is particularly relevant to
this scoping project concerns the relationship between scores derived from different tests.

In the Hale et al. report there is mention of equivalence between a score on the screening test and
levels of performance on other recognized English language qualifications (such as a score on IELTS,
mentioned above). There will therefore need to be a formal mechanism to ensure that the minimum
threshold/s set on the selected screening test is/are at least approximately equivalent to the
minimum requirement/s on any other test recognized as equivalent. To achieve this some evidence
will first need to be produced showing some relationship between the two measures with respect to
the abilities they claim to measure. Then, if this relationship can be demonstrated, recognized test
equating procedures (Holland & Dorans, 2006) should be undertaken to establish how the relevant
cutscore/s (or minimum standard/s) on one scoring scale can be equated to scores on the other
measure. The same would ideally apply to the proposal (in Hale et al. 2012, pp. 10-11) that
completing a particular English language course might be considered equivalent to a given level on
the selected language screening test or tests. The mechanism for such cross-test equating or for
ensuring the comparability of exit English proficiency standards for particular language courses will
be considered later in this report.

Utility
An assessment may be valid, and reliable, but high cost or unreasonably elaborate
procedures may prevent its use. The utility of an instrument is an indication of how practical
it is to use in a given situation.

Although the proposed English testing procedures are to be voluntary, it is important that practical
considerations do not act as a major deterrent to candidates undergoing these procedures. Since the
test is designed for preliminary screening purposes, one such consideration would be that the
screening can be completed in a reasonably short period of time. It is also necessary that online
delivery mechanisms are not unduly cumbersome and that score reports are delivered quickly and in
a form that is readily intelligible to users. It will also be important that issues surrounding the
development and/or implementation of the selected test/tests do not entail prohibitive costs either
for NAATI or for the candidates themselves.

Worth noting here is a tension between the requirement of utility and that of validity as outlined
above. A truly comprehensive testing procedure that samples from all components and registers of
language ability of possible relevance to translating and interpreting would certainly be desirable.
However, given the need for cost-effectiveness and for efficiency and immediacy of scoring, it may
be necessary to select a test that assesses language proficiency indirectly, rather than adopt a direct,
performance-based approach. For example, speaking and writing tasks that require scoring by
trained human raters, rather than by a computer engine, are both time consuming and costly to
administer and cannot by their very nature offer immediate feedback to test candidates. And while
the science of automated scoring of speaking and writing is rapidly advancing, developing tools
which yield accurate estimates of performance on open-ended tasks requires a level of resources

11
and expertise which are only available to large commercial testing agencies whose systems are strictly proprietary. For practical reasons therefore, it may be necessary to assess at least some components of language ability via indirect items that bear little obvious resemblance to real world communicative tasks. It may also be necessary to restrict the range of skills covered on the test. Feedback on performance may accordingly be limited, providing only a broad estimate of ability, rather than detailed diagnostic feedback. Given that the stakes associated with preliminary screening are relatively low, in the sense that the tool is for preliminary self-assessment purposes only and does not itself determine readiness to practice as an interpreter or translator, such compromises may be worth considering.

**Equity**

Equity is the principle that recognizes and accommodates to differences in preparedness among individuals and groups, because of differences in familiarity or exposure to the tasks of interest or to the modality of testing.

Acknowledging group or individual differences is a question of test fairness. Any accommodations to particular groups would be for the purpose of ensuring that all candidates, regardless of their background or circumstances, have equal chances to display their true level of ability on the test. Most testing agencies have some kind of sensitivity review at the test design stage so that items that may be disturbing or cause offence to particular groups are not included on the test. Some also conduct statistical bias analyses (Elder, 2012) to ensure that construct irrelevant factors (those that are irrelevant to the ability under test) are not interfering unduly with test scores, although such analysis tend to be undertaken only with high stakes tests used for admission to academic institutions or for professional certification.

In our case, although the proposed screening test is relatively low stakes, fair treatment remains important. It will be necessary to ensure that any English test that is proposed for the intended self-screening purpose is readily accessible to candidates seeking T&I accreditation in Australia, regardless of where they are located, and that appropriate information about the test and its purpose, content and format is provided in a manner that is understandable to all candidates, regardless of level of education or language background concerned (please refer to the ILTA Guidelines for Practice [http://www.ultaonline.com/index.php/enUS/component/content/article?id=122] for information on the responsibility of testing agencies to provide comprehensive and accurate information to stakeholders). Practice test tasks, to ensure that candidates understand the format and can manage the online delivery mode before commencing the test, should be available online. This is particularly important when the selected test measures language ability via tasks or item types which are unfamiliar to test-takers. Accommodations involving individualized test conditions may also need to be considered for candidates with disabilities, provided that any such accommodations do not alter the construct under test, thereby distorting the meaning of test scores (Abedi, 2012)². Reporting of performance levels should also be worded as simply and transparently as possible. To ensure that the above conditions are being met to the satisfaction of users, it would be important to gather feedback following an initial test pilot with members of the target population as mentioned later in this report.

---

² The provision of accommodations such as an aural-oral version of a written test for visually impaired candidates as proposed in the Hale et al. report (page 40) may for example be problematic if the purpose of the test is to measure the construct of writing ability.
Finally, it is critical that a test practice of preparing fruitlessly for a test is a negative consequence that is termed negative test washback in testing circles (Wall, 2012). It is important that test takers receive appropriate information about the limitations of any test that is selected, and that, to limit any negative washback, advice about the kinds of activity needed to improve their language skills is provided, as proposed in the Hale et al. (2012) report.
3. Methodology

The above principles were used as a guide in our search for possible test suitable for screening purposes.

The search involved a scan of the databases of all major testing agencies and clearing houses as well as reviews of language tests in the key language testing and applied linguistic journals (such as Language Testing and Language Assessment Quarterly). A request for information was also sent to the Language Testers’ professional list L-test-l, to ensure that no relevant test was excluded from consideration.

The tests identified from this search were reviewed by the researchers in terms of their fitness for the intended purpose along with any available evidence for validity, reliability and utility. All tests that were not delivered online and did not include automated scoring for at least some of their components were eliminated from the final shortlist along with those for which there was no available validity or reliability evidence in the form of technical reports or published research. Twelve tests were identified for the shortlist and reviewed. The available information on these shortlisted tests was carefully scrutinized and practice tests were taken by one or other of the researchers to better understand the test format and to identify any obvious issues with the test delivery platform.

4. Findings

Summary information about each shortlisted test is presented sequentially (in alphabetical order) in Appendix C below in tabular form followed by a more detailed prose description of its qualities. A sample score report for each test showing the information provided to test-takers is shown in Appendix D.

The test descriptions raise a number of issues that will need to be considered by NAATI in deciding on the best option for pre-examination English proficiency screening. Most of these issues were raised in the Statement of Requirements above but will be revisited in our discussion below in relation to the particular tests reviewed.

5. Discussion

Validity

Most of the short listed tests claim to cater for a broad range of abilities, ranging from near beginner to very advanced, but on closer inspection this is not always the case. Both Aptis and Password currently cater only for learners in the IELTS 3.5-6.5 range and do not discriminate at higher ability levels. The same is true of the Versant English Placement test, which is a condensed version of the Versant English Speaking and Versant Writing and does not provide any information at the top end of the Pearson Global English scale used for reporting performance. As noted in our review of these tests, the relative “easiness” of these tests limits their value for the intended self-screening purpose given that NAATI may wish to set its recommended thresholds higher than 6.5 particularly for those intending to sit the more advanced T&I exams.

As foreshadowed in our Statement of Requirements, many online tests measure language proficiency indirectly and do not include a speaking and writing proficiency component. This may limit their predictive power to some extent, although most of these tests report moderately strong correlations with other more direct, performance-based measures. Included in this group of indirect tests are the OOPT, AEST, DELNA Screening, CaMLA and the EFSET. Dialang claims to assess writing,
but, as noted in our review (please see Appendix C), does so very indirectly, via multiple choice items involving minimal text production. It should be noted that DELNA Screening and AEST include a writing component as an optional extra, but these components are not automatically scored and add to the overall cost of the procedure.

All of the tests (with the exception of EFSET) measure vocabulary and grammar knowledge, along with other receptive skills in some cases. Two, AEST and DELNA Screening, do not assess listening skills, which may be of particular importance for interpreters. This may be a consideration in NAATI’s choice of an appropriate instrument, although there is evidence that overall score on both tests does correlate with listening skills on another professionally validated English measure. Both AEST and DELNA Screening on the other hand include an extended reading test, which the test taker must scan and parse under time pressure. This kind of textual competence is not measured by the other tests in this group, which rely on relatively brief input texts.

Although a number of tests include a speaking and/or writing component, administered either concurrently with or in addition to other test modules, only four (Progress, Versant English Speaking, Versant Writing and the Versant English Placement test) provide for automatic scoring of these skills. These tests, all in the Pearson group, use sophisticated automated scoring engines developed after years of careful research. The Progress test, according to its designers, is however designed as a measure of achievement to measure progress in the context of a particular language instruction program rather than as a standalone proficiency measure like the other three. The remaining standalone tests in the Pearson suite provide evidence of reliable estimates of ability yielding scores on a par with those assigned by human raters and correlate strongly or moderately with other more communicatively oriented measures of the same skill. The fact that speaking scores are assigned based on monologic performance on artificial tasks such as repeating or reconstructing sentences, rather than through the kind of dialogic interaction with a human interlocutor that is required in the interpreting profession, is a validity limitation that should perhaps be tolerated on practicality grounds, although it may be perceived by test takers as artificial.

Reliability
All tests reviewed met acceptable standards of reliability and therefore can be seen as providing stable estimates of ability with the exception perhaps of Dialang, where the test taker’s control of the order of the tests taken, the time spent on each item and whether feedback is provided on an item by item basis or at the end of each test section may produce unacceptable levels of variability in score meaning. While this may not be a major concern for individual self-diagnosis, it makes it difficult to set a minimum threshold on the test that will accurately estimate the readiness of the whole cohort to take the NAATI exam. There are also limited number of test versions making it quite likely that someone who resits the test will encounter the same version twice. This is less likely with the other tests reviewed, all of which have several parallel forms or draw randomly on a large item bank and do not, like Dialang, provide feedback on correct or incorrect answers.

Comparability
All tests other than AEST and DELNA, which were developed for local use, have benchmarked their scores to CEFR levels and provided evidence of the benchmarking process. However, as already noted above, such levels, in spite of their international currency, are somewhat vague and broad in scope. Saying that a candidate is at the B2 level could mean that their IELTS scores might be as low as 5 and as high as 6.5. For this reason it is not unlikely that many candidates who retest after a year may remain at the same level on the CEFR, even though their language ability might have improved considerably. This would certainly be discouraging. It also creates problems with regard to equivalences. How could CEFR levels be equated to current NAATI requirements? Although the score of IELTS 7.5 that is currently required for overseas-trained applicants for a NAATI qualification above the professional level falls squarely within the CEFR C1 range, not all candidates classified as C1
would be able to achieve 7.5 on IELTS. A finer scale would therefore seem desirable. The Versant speaking test score has been equated to a score range on the TOEFL and the TOEIC (which could in turn be linked to IELTS via a score comparison tool (https://www.ets.org/toefl/institutions/scores/compare/). An AEST or DELNA score can likewise be equated to a bandscore on DELA (which could in turn be linked loosely to IELTS based on data already available). While such equivalences will always be approximate, given that all the tests differ in the way they construe language proficiency, they are nevertheless likely to be more precise than any link to the much broader CEFR levels.

Utility
Practical issues surrounding the use of the various tests including time, delivery, pricing and quality of delivery are considered below.

Time
The time allowance for each test varies widely and is related to the question of construct/skill coverage. The tests which take longer than an hour (Aptis, Dialang, EFSET (Plus) and DIALANG) assess a broader range of skills. Two exceptions are the Versant English Placement test which assesses reading, listening, speaking and writing in 50 minutes and the Progress test which assess all four macroskills as well as vocabulary and reading in 40-55 minutes. However neither test taps language abilities at the higher levels needed for the T&I profession.

Delivery
An important issue for the individual test-takers is that taking the test online is straightforward and does not present technical difficulties. Most of the test websites indicate system requirements for successful delivery of the relevant test. EFSET, taken by two researchers on a Mac computer, presented difficulties in this regard.

While all the shortlisted tests are administered online, the manner in which they are delivered to candidates varies. Some testing sites can be accessed directly by candidates from the internet via a link provided either on the webpage or sent via email to the registering candidate (Dialang, EFSET, Versant tests). The others are site-licensed to institutions who themselves take responsibility for registering or allocating places to individual test takers and collecting their scores. Some tests like Versant are available in either individual or group delivery mode but only a single form of the test is made available to individuals, meaning that resitting the test may produce invalid results.

It should be noted that most testing agencies recommend that their test be administered under proctored conditions to ensure that the candidates’ identity can be verified and test security is not jeopardized in ways that could threaten the validity of inferences drawn from test scores. This would however limit some test takers’ access to the test preventing them from taking it at a time and place of their choosing. If the test is to be voluntary, there is no obvious reason why candidates would attempt to cheat or artificially boost their scores and it could be argued that invigilating the test potentially raises the stakes and means that some test takers may form the perception that satisfactory performance on the screening test is a necessary pre-condition for taking the NAATI exam, rather than simply an indication of their preparedness for individual decision-making purposes.

NAATI will therefore need to consider the pros and cons of allowing candidates free access to the selected test from a time and place of their choosing, where this is an option provided by the testing agency. One possibility would for NAATI to host the relevant delivery platform and issue download details to each candidate without invigilating the tests. Scores could be provided to both the
candidates for decision-making and to NAATI for research purposes as discussed in the concluding section of this report.

**Pricing**
The shortlisted tests vary in cost. Those which are free to users (Dialang and EFSET) may be problematic as noted above because of lack of technical support. Other tests range in cost from as low as Aus$4.00 per head (AEST) to as high as $77 (Aptis - all components). A number of the tests are priced according to the number of test components included in the package (AEST, DELNA Screening, Aptis, Password). Our communications with the various testing agencies, suggest that these costs could either be paid directly by test takers to the agency concerned or, if outlaid by NAATI for a site-licence, recouped from individual test-takers when they apply to take the test. While reasonable costs should be relatively easy to justify to test takers if their purpose of signalling to a candidate that s/he may not be ready to take the much more expensive NAATI exam is made clear, the more expensive test options may be perceived by some as burdensome and may dissuade them from undertaking the self-assessment.

**Quality of feedback**
The tests also vary in the amount of feedback provided, with the Versant tests at one end of the spectrum offering both a score and descriptive feedback on each dimension of ability tested, and advice on English development strategies, and, at the other end, the AEST and DELNA Screening which offer a three-way classification indicating whether the test performance is clearly above or below a particular level\(^3\) or on the borderline. Of the remaining tests CaMLA, designed primarily for course placement purposes, provides only an overall score whereas Aptis, Progress, EFSET and Password report both a score for overall performance and a breakdown by skill and or knowledge area. The Oxford Placement test reports both a combined score and a separate score for each of its two components (Listening and Structure). Provided that there is empirical evidence supporting the separation of different skill/knowledge components\(^4\), the provision of more detailed information is a bonus, and may make the test experience more useful and appealing for candidates.

**Equity**
All the tests reviewed offer information to users, either via a website and or in the form of handbooks, short videos with sample items, practice items/test forms. Two tests, the Oxford Placement Test and the Versant Placement Test, offer information on their score report about factors that might influence test performance either positively or negatively (time taken to complete the test in the former case, and typing speed and accuracy in the latter). However it is not really clear how such information would be used for the current purpose of giving test takers themselves the opportunity to get immediate feedback on their ability without any institutional intervention. Since most of the test reviewed are placement tests, rather than high stakes selection tools, there are understandably no accommodations for candidates with disabilities. Provision for individual difference is perhaps best dealt with by NAATI in the form of general caveats warning test takers of factors which might diminish or augment test performance for reasons independent of their true English language ability and which might accordingly limit the value of the test score as an indicator of readiness to take the NAATI exam.

---

\(^3\) The level can be adjusted according to context and test purpose.

\(^4\) Empirical support for skills separation involves a statistical investigation of test dimensionality, which has been reported in relation to a number of the reviewed tests.
6. Conclusion and recommendations

The above discussion suggests that a number of tests can be removed outright from our shortlist. One such test is Dialang because of how test conditions can be manipulated by test candidates to the possibly detriment of accurate scores. Password, Aptis (in its current form) and the Versant English Placement test also appear unsuitable for the current purpose on the grounds they span too limited an ability range for the relevant population. Progress can also be excluded because it is designed to measure course achievement rather proficiency. EFSET is also best avoided on the grounds of technical problems encountered by two researchers attempting the test on more than one occasion\(^5\). Of the others both the AEST and DELNA Screening, while limited in skill coverage, are highly economical and work well as predictors of proficiency in other skills. The AEST at circa $4.00 per unit is even more economical than DELNA Screening at $8.00. Moreover the C-test task used in the AEST draws on a broader range of abilities than the corresponding vocabulary task in the DELNA Screening, requiring active text completion rather than just passive recognition of word meaning. Neither of these tests have a listening or a speaking component, which is a serious limitation particularly for those seeking certification as interpreters.

The Versant English Speaking Test seems the most suitable option for online assessment of both listening and speaking ability at the levels of ability likely to be relevant for the interpreter role. The listening input is somewhat limited but correlates quite well with listening proficiency as measured by high stakes language admissions tests. Speaking proficiency as noted in the test description is measured indirectly via rather artificial tasks but again, according to available evidence, functions well as a surrogate measure of this skill. It meets the requirement of efficiency in that is completed on line, takes only 15 minutes to complete and offers immediate feedback to test takers. This test also has the advantage of a detailed score report together with advice on actions that candidates can take to improve their English and clear links to not only the CEFR, but also to scores on other recognized proficiency tests. The latter could be useful in linking scores on this test with those on recognized tests such as IELTS, which are currently required for NAATI accreditation purposes. A further characteristic worth highlighting is that this test can be taken by individual test takers. On the other hand this comes at a price as only one test form is available to individual test takers, whereas those registering via an institutional licence can access the full item bank allowing for different (but parallel) items to be presented to test takers who might need to resit the test. At Aus$46.00 per capita (based on current exchange rates) the cost of taking the Speaking test is considerable but perhaps not prohibitive.

As for the writing skill, crucial for translating, it appears that there are two possible options: one would be to include the available option of a computer delivered AEST or DELNA writing task which would require human scoring. Including this option, it should be noted, would add considerably to the unit cost of these tests. The other is the computer scored Versant Writing which has strengths (self-access, concurrent assessment of reading skills, detailed score report) and weaknesses (somewhat artificial tasks) similar to the Versant English Speaking but is more expensive at Aus$56.00 per unit. A valuable service offered in relation to both these Versant tests is the provision for local standard-setting, using procedures and materials provided by Pearson for this purpose (see further discussion below).

Should NAATI consider the cost of the Versant tests to be a major obstacle, it may wish to consider advising candidates to sit for a cheaper test such as the AEST or DELNA screening as a means of determining their general proficiency level and deciding whether it is worth their while sitting the more expensive tests which offer a profile of their speaking and/or writing skills. This would have the

\(^5\) It is quite possible that such problems will be resolved in time, as EFSET is a newcomer to the testing scene, so this option might be reconsidered in due course.
advantage of signalling the need for general proficiency as well as skill-specific competences. Those placed in the “At Risk” category (to be set at a suitable level for the NAATI test taker population) could be advised not to take the Versant English (Speaking) or Writing tests until they had done further English study. The mechanisms for delivering this two-tier process would need to be considered further and the thinking behind it carefully explained to NAATI applicants. One possibility would be for NAATI to purchase a site licence for AEST or DELNA Screening and then simply direct candidates scoring at or above the minimum required level to the self-access versions of Versant English Speaking and/or Versant Writing as a means of getting further information about their productive skills should they so desire.

Another option, mentioned in the Hale et al. (2012) report, would be for NAATI to commission development of a custom-built proficiency test with specifications that would be common across English and the other languages serviced by the organization and that would specifically target the needs of the T&I profession and include a speaking and writing component. Developing and validating such a test would take at least two years and would require substantial resources. Given that there are viable and ultimately cost-neutral (for NAATI) alternatives, contemplating such an option does not seem warranted.

An alternative which would perhaps be more worthy of consideration would be to develop an additional computer-scored listening task for the AEST, which as noted above, in its current form contains only written input texts. (The Language Testing Research Centre has already developed a prototype listening task as part of its custom-built placement tests in a range of other languages, which would lend itself well for this purpose.) Developing trialling and validating this additional task on would also take considerable time and money, but would be less costly than developing an entire new test. The benefit would be that AEST might then function as a standalone screening procedure, if it proved be sufficiently predictive of spoken proficiency to be used without the Versant English test.

In any case, whichever option NAATI pursues, it will be important, as foreshadowed at the beginning of this report, to establish or verify the equivalence between the scoring scale of the selected test and the scoring scale of other standardized tests like IELTS and TOEFL used for admission to T&I training courses. This kind of equating study could be done in the context of an invigilated trial of the new test with beginning T&I students who have recently sat one or other of these admissions tests and are willing to report their scores. The trial could also be set up to elicit user feedback on the utility of the chosen test/tests and associated score report.

A standard-setting exercise will also be needed to establish a minimum threshold, or series of thresholds at or above which test candidates can be deemed fit to sit one or other of the NAATI examinations. This will necessitate the following steps:

1. Convening a panel of experts in interpreting and translating to consider the kind and level of language skills and knowledge required for a candidate to be considered linguistically ready to sit the NAATI examinations. The panel would comprise both academics responsible for teaching the theory and practice of interpreting and translation and qualified practitioners with experience of T&I work in a range of contexts and purposes.

2. Conducting a workshop or series of workshops involving:
   o Briefing of the panel about the meaning and purpose of standard setting
   o Consideration by the panel of required language levels in light of work done on developing standards for the profession, referred to at the outset of this report, of

Reviewing available test in other languages will be the focus of Part II of this report.
existing level statements such as the CEFR and experience of teaching and/or assessing T&I students;

- Familiarizing the panel with tasks and items or sample materials from the selected test;
- Eliciting judgements from panel members on minimum cut-score/s on each test section following recognized standard setting procedures;
- Determination of appropriate cut-scores through statistical analysis;
- Preparation of a report describing processes and outcomes of the workshop.

Once cut-scores are set it would desirable to verify these empirically in a subsequent study comparing the test performance of a representative sample of test takers with their subsequent performance on the NAATI examinations\(^7\) and by administering the language test to students on completion of certified interpreter or translator courses, whose average proficiency level would be expected to be located at or around the established cut-score. A research proposal for this purpose could be developed by NAATI in collaboration with relevant researchers.

Such steps will be necessary to ensure that the selected test is be appropriately aligned with current requirements for professional accreditation and that recommended minimum thresholds are defensible and in accordance with the values and expectations of the T&I profession.

Finally we would emphasize NAATI’s responsibility, in line with current standards for good practice in language testing, to clearly inform users about its English proficiency policy, highlighting the importance of English proficiency more generally as a means for operating effectively credibly in the professional T&I domain and justifying the purpose and intended use of the selected English test or tests. Any test that is recommended for screening purposes should be accompanied by clear advice for users about its content and format and the meaning of test scores including any caveats on their interpretation. Avenues for English language development for those candidates whose performance indicates lack of readiness to sit the NAATI exams should also be outlined.

In summary, we offer a number of specific recommendations as detailed below. The first two relate to the choice of a suitable test and associated framework of reference, the second and third relate to the research that needs to be conducted in relation to this test and the remaining four to the policy context for the test and the way in which it is used.

We recommend:

1. That NAATI consult with T&I experts to consider available language proficiency frameworks (such as CEFR) and language testing regimes (such as IELTS) and determine what level or levels they deem most suitable as an English proficiency benchmark for the T&I profession;

2. That NAATI and T&I experts to consider the following test options for English proficiency screening as discussed above;
   - Option A: **Versant English Speaking** for all applicants for the interpreting exams and **Versant Writing** for all applicants for translating.
   - Option B: **AEST** (or **DELNA Screening**) for all NAATI applicants followed by a recommendation to sit **Versant Speaking** (for interpreters) or **Versant Writing** (for translators) for those who score above a given threshold.
   - Option C: A customised version of **AEST** which includes a listening component to be used with or without the Versant add-ons.

---

\(^7\) Ideally such a study would involve candidates whose scores on the English language test ranged both above and below the recommended cut scores.
3. That NAATI negotiate with the relevant agency/ies to run a year-long trial of the chosen option with both T&I students and applicants for the NAATI exam for whom English is a second language.

4. That NAATI conduct a standard-setting exercise with members of the T&I profession to establish recommended minimum score thresholds on the chosen test or tests above which candidates might be considered ready to sit the relevant exam;

5. That NAATI develop a language proficiency policy (in relation to English and other languages) which sets out language proficiency expectations for entry and exit from accredited T&I courses and explains the rationale for preliminary language proficiency screening as a means of ascertaining readiness to take the NAATI exams, as well as the nature of the chosen test or tests and how scores should be interpreted and acted on;

6. That students register voluntarily with NAATI to take the chosen test or tests for a fee (consistent with the costs of NAATI's site-licensing arrangements with the relevant agency);

7. That immediate feedback on test performance be provided to candidates where possible and used by NAATI for research purposes only;

8. That the English screening program, once implemented, be reviewed on a regular basis to enable adjustments to policy and practice in light of new developments in the language testing field.

7. References


Elder, C., & Knoch, U. & Zhang, R. Diagnosing the support needs of second language writers: Does the time allowance matter? TESOL Quarterly 43,2: 351-360


### Appendix A. Entry requirements for accredited translator and interpreter courses in Australia

<table>
<thead>
<tr>
<th>Institution</th>
<th>Contact Information</th>
<th>Entry Requirements</th>
</tr>
</thead>
</table>
| **Abbey College Australia**                      | Contact: 2 9212 4470 or info@abbeycollege.edu.au                                   | • Advanced Diploma of Translating  
IELTS 6.5 or completion of a recognised degree or diploma or equivalent in any discipline from an English speaking country |
| **Australian Ideal College**                     | Contact: student Support Officer on 8833 3604 or den@deaf.nsw.edu.au                | • Advanced Diploma of Translating  
IELTS 6.0 overall; or complete a recognized Degree or Diploma or equivalent in any discipline from an English speaking country; or undertake an entry test on both English proficiency and translation of two passages to demonstrate adequate competency in both English and LOTE as part of the enrolment process.  
• Diploma of Interpreting  
IELTS 5.5 overall with min. 6 in Speaking and Listening subtests; or an upper intermediate level certificate by having successfully completed an English program within an English speaking countries; or an entry test; or equivalent with the level of paraprofessional interpreter to demonstrate adequate competency in both English and LOTE as part of the enrolment process. |
| **Macquarie University**                         | Contact: 02 9850 7102 (Postgraduate) 9898 (NAATI Accreditation)                    | • Master of Interpreting  
• Master of Translating  
• Graduate Diploma of Translating and Interpreting  
IELTS of 6.5 overall with min. 6.0 in each band, or equivalent  
• Master of Advanced Translation and Interpreting Studies  
• Master of Translating and Interpreting Studies  
IELTS of 7.0 overall with min. 6.5 in each band, or equivalent |
| **Multilink Academy of Interpreting and Translating (MAIT)** | Contact: 02 9261 8885 or info@mait.edu.au                                         | • Advanced Diploma of Translating  
IELTS 6.5 or above or proof of work experience in translating |
| **Sydney Institute of Interpreting and Translating (Sydney)** | Contact: 02 8090 3266 or info@siit.nsw.edu.au                                    | • Advanced Diploma of Interpreting  
IELTS 6.5 or equivalent; or a min. score of 70 at the SIIT entry test  
• Advanced Diploma of Translating  
IELTS 6.5 in all subtests  
• Diploma of Interpreting- Mandarin, Cantonese, Nepali, Punjabi, Arabic, Hindi, Korean  
IELTS 5.5 with min. 6 in Speaking and Listening subtests; the speaking and listening; or upper intermediate level or equivalent in EAP program; or be able to demonstrate vocational proficiency through an entry examination (50% or higher) |
<table>
<thead>
<tr>
<th>No.</th>
<th>Institution</th>
<th>Contact Information</th>
<th>Qualifications</th>
<th>Pathways into the Qualification</th>
</tr>
</thead>
</table>
| 6   | Sydney TAFE | Contact: 1200 360 601 or 02 9217 4106 | • Advanced Diploma of Translation  
• Advanced Diploma of Interpreting  
• Diploma of Interpreting | No details are available. On the website, it says “There are no entry requirements for the qualifications”. Pathways into the qualification:  
Candidates may enter the qualification through demonstrated language proficiency in English and another language at a level required to undertake and complete the requirements of training and assessment for this qualification. Proficiency may be demonstrated through recognised or accredited language qualifications or through an RTO determined assessment of current competency. |
| 7   | TAFE NSW- Northern Sydney Institute | Contact: 131 674 | • Advanced Diploma of Interpreting  
• Advanced Diploma of Translating  
• Diploma of Interpreting | No details are available. On the website, it says “There are no entry requirements for the qualifications”. Pathways into the qualification:  
Candidates may enter the qualification through demonstrated language proficiency in English and Mandarin at a level required to undertake and complete the requirements of training and assessment for this qualification. Proficiency may be demonstrated through recognised or accredited language qualifications or through a Registered Training Organisation determined assessment of current competency. |
| 8   | TAFE NSW- South Western Sydney Institute | Contact: 13 7974 or swsi.customerservice@det.nsw.edu.au | • Diploma of Interpreting  
IELTS 5.5; or TOEFL iBT 46; or TAFE NSW Certificate 3 in Advanced English for Further Study |
| 9 | UNSW Australia  
Contact: UNSW Arts & Sciences  
02 9385 3697  
| 10 | Wentworth Institute  
Contact: 02 8252 9999 or  
info@win.edu.au  
| 11 | Western Sydney University  
Contact: Dr Uldis Ozolins,  
Director of Languages,  
Interpreting & Translation,  
TESOL programs  
02 9772 6668 or  
U.Ozolins@westernsydney.edu.au  
| 12 | Harvest Education Technical College  
Contact: 07 3344 1922 or  
sunnybank@hetc.edu.au  

<table>
<thead>
<tr>
<th>Program</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor of Arts (Interpreting and Translation)</td>
<td>IELTS 7.0 with min. 6.0 in each subtest</td>
</tr>
<tr>
<td>Graduate Diploma in Interpreting</td>
<td>TOEFL CBT 250 with min. 5.0 in writing</td>
</tr>
<tr>
<td>Graduate Diploma in Translation</td>
<td>TOEFL iBT 100 with min. 21 in Writing and 18 in all other subtests</td>
</tr>
<tr>
<td>Master of Interpreting and Translation</td>
<td></td>
</tr>
<tr>
<td>Master of Translation and TESOL</td>
<td></td>
</tr>
<tr>
<td>Diploma of Interpreting</td>
<td>IELTS 5.5 with min. 6 in Speaking and Listening; upper intermediate level or equivalent in the EAP program; or be able to demonstrate vocational proficiency through an entry examination</td>
</tr>
<tr>
<td>Advanced Diploma of Translating</td>
<td>IELTS 7.0; or pass a written entry test designed to assess the level of reading, comprehension and writing proficiency in both languages; or complete the Diploma of Translation with a score of at least 70% in Translation; or NAATI accreditation at the paraprofessional level in the relevant language and proven work experience; and a native-like proficiency in a relevant LOTE language.</td>
</tr>
<tr>
<td>Diploma of Interpreting</td>
<td>IELTS 6.5 overall with min. 6.0 in each subtest</td>
</tr>
<tr>
<td>TOEFL iBT 90 overall with min. 23 in Writing, 22 in Reading, Listening and Speaking</td>
<td></td>
</tr>
<tr>
<td>Pearson 64 overall with min. 54 in each subtest</td>
<td></td>
</tr>
<tr>
<td>Successful completion of the University English Entry Course with a min. grade of C+ and a min. score of 20 in Writing</td>
<td></td>
</tr>
<tr>
<td>Cambridge Certificate in Advanced English 176 overall with min. 169 in each subtest Cambridge Certificate of Proficiency in English 180 overall with min. 180 in each subtest Successful completion of ANU College Access Program with a min. of 70% overall</td>
<td></td>
</tr>
<tr>
<td>Diploma of Interpreting</td>
<td>IELTS 6.5 overall with min. 6.0 in each subtest</td>
</tr>
<tr>
<td>TOEFL iBT 90 overall with min. 23 in Writing, 22 in Reading, Listening and Speaking</td>
<td></td>
</tr>
<tr>
<td>Pearson 64 overall with min. 54 in each subtest</td>
<td></td>
</tr>
<tr>
<td>IELTS 7.0 with min. 6.0 in each subtest</td>
<td></td>
</tr>
<tr>
<td>TOEFL iBT 90 overall with min. 23 in Writing, 22 in Reading, Listening and Speaking</td>
<td></td>
</tr>
<tr>
<td>TOEFL iBT 100 with min. 21 in Writing and 18 in all other subtests</td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>Contact</td>
</tr>
<tr>
<td>-------------</td>
<td>---------</td>
</tr>
</tbody>
</table>
| Sydney Institute of Interpreting and Translating (Brisbane) | 07 3088 2850 or info@siit.nsw.edu.au | • Advanced Diploma Translating (IELTS 6.5 in all macro skills; Australian Core Skills Framework at or around level 4 across learning, reading, writing, oral communication core skills; entry selection by a pre-interview language task and interview)  
• Diploma of Interpreting (IELTS result of 5.5 with a min. score of 6 in the speaking and listening; or upper intermediate level or equivalent in EAP program; or be able to demonstrate vocational proficiency through an entry examination (50% or higher)) |
| University of Queensland | Mrs Sharon Crossman, Student & Academic Admin. Officer 07 336 52013 or s.crossman@uq.edu.au | • Graduate Diploma in Chinese Translation and Interpreting  
• Master of Arts in Chinese Translation and Interpreting  
• Master of Arts in Chinese Translation and Interpreting (Advanced)  
• Master of Arts in Japanese Interpreting and Translation (IELTS 6.5 with min. 6 in all subtests) |
| Australian Institute of Translation and Interpretation (Adelaide) | 03 9620 1618 or info@aiti.edu.au | • Advanced Diploma of Translating (IELTS score of 6.5 or (see below))  
• Diploma of Interpreting - Mandarin  
An IELTS score of 6.0 or (see below)  
; completion of a recognized Degree or Diploma or equivalent in any discipline from an English speaking country. In addition, all students are required to undertake an intake test to demonstrate adequate competency in both English and LOTE as part of the enrolment process. |
| TAFESA | 1800 882 661 | • Advanced Diploma of Translating  
• Diploma of Interpreting  
Written and oral proficiency assessment in English and another language*  
*Based on the information on the web, these courses are not offered to international students. |
| University of Adelaide | 08 8313 7335 | • Graduate Diploma in Translation and Transcultural Communication  
• Master of Arts (Translation and Transcultural Communication)  
IELTS 6.5 overall with min. 6 in all subtests |
| Australian Institute of Translation and Interpretation (Melbourne) | 08 7088 2778 or info@aiti.edu.au | • Advanced Diploma of Translating (IELTS 6.5 or (see below))  
• Diploma of Interpreting (IELTS 6.0 or (see below))  
; completion of a recognized Degree or Diploma or equivalent in any discipline from an English speaking country. In addition, all students are required to undertake an intake test to demonstrate adequate competency in both English and LOTE as part of the enrolment process. |
<table>
<thead>
<tr>
<th>Institution</th>
<th>Contact</th>
<th>Program Details</th>
</tr>
</thead>
</table>
| Australian Institute of Translation and Interpretation (Queens Road, VIC) | 08 7088 2778 or info@aiti.edu.au | Diploma of Interpreting  
IELTS 6.0 or  
; completion of a recognized Degree or Diploma or equivalent in any discipline from an English speaking country. In addition, all students are required to undertake an intake test to demonstrate adequate competency in both English and LOTE as part of the enrolment process. |
| Cambridge International College | 03 9663 4933 or info@cambridgecollege.com.au | Advanced Diploma of Translating  
IELTS 6 with min. 6 in all subtests |
| Monash University | Mr Marc Orlando, Coordinator of Master of Interpreting and Translation Studies  
03 990 52252 or Marc.Orlando@monash.edu | Master of Interpreting and Translation Studies  
IELTS 6.5 with min. 6 in all subtests  
TOEFL iBT 79 with min. 21 in Writing, 12 in Listening, 13 in Reading, 18 in Speaking  
Pearson Test of English 58 with min. 50 in all subtests  
Cambridge Advanced and Proficiency 176 with min. 169 in all subtests  
Master of Interpreting and Translation Studies  
IELTS 6.5 with min. 6 in all subtests  
TOEFL iBT 79 with min. 21 in Writing, 12 in Speaking, 13 in Reading, 13 in Listening  
Pearson Test of English 58 with min. 50 in all subtests  
Cambridge English Advanced Grade B with min. Good in all subtests  
Diploma of Interpreting  
IELTS 6.0 with min. 5.5 in all subtests  
TOEFL iBT 79 with min. 19 in all subtests  
Pearson Test of English 50 with min. 42 in all subtests  
Cambridge English Advanced Grade C with min. Borderline in all subtests |
| Universi of Melbourne | 13 63 52 or arts-gradadmissions@unimelb.edu.au | Master of Translation  
IELTS 6.5 with min. 6 in all subtests  
TOEFL iBT 79 with min. 21 in Writing, 18 in Speaking, 13 in Reading, 13 in Listening  
Pearson Test of English 58 with min. 50 in all subtests  
Cambridge Advanced and Proficiency 176 with min. 169 in all subtests  
Master in Translation Studies  
IELTS 6.5 with min. 6 in all subtests  
TOEFL iBT 82 with min. 22 in Writing, 20 in Speaking, 18 in Reading, 20 in Listening  
PTE 64 with min. 59 in Speaking and Writing, 54 in Listening and Reading subtests  
Cambridge Advanced 58 (B grade) or above; Proficiency 45 (C pass) or above |
### Appendix B. Global scale descriptors for Common European Framework of Reference (CEFR) levels

<table>
<thead>
<tr>
<th>Level</th>
<th>Proficiency</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>Proficient User</td>
<td>Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.</td>
</tr>
<tr>
<td>C1</td>
<td>Independent User</td>
<td>Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.</td>
</tr>
<tr>
<td>B2</td>
<td>Independent User</td>
<td>Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.</td>
</tr>
<tr>
<td>B1</td>
<td>Independent User</td>
<td>Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics, which are familiar, or of personal interest. Can describe experiences and events, dreams, hopes &amp; ambitions and briefly give reasons and explanations for opinions and plans.</td>
</tr>
<tr>
<td>A2</td>
<td>Basic User</td>
<td>Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.</td>
</tr>
<tr>
<td>A1</td>
<td>Basic User</td>
<td>Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.</td>
</tr>
</tbody>
</table>
Appendix C. Findings of shortlisted tests

Summary information about each shortlisted test is presented sequentially (in alphabetical order) below in tabular form followed by a more detailed prose description of its qualities. The headings on the vertical access of each table are linked to the principles identified above. Rows 1 indicates the main use for the test. Rows 2 – 5 cover issues pertaining to Validity, Rows 6 and 7 with Reliability, Row 8 with Comparability, Rows 9-11 with Utility and Row 12 with Equity.

1. Academic English Screening Test/Post-entry Assessment of Academic Language

Table 1. Academic English Screening Test/Post-entry Assessment of Academic Language
Contact: Dr Ute Knoch, Language Testing Research Centre Director, University of Melbourne. +61 3 8344 5206 or uknoch@unimelb.edu.au

<table>
<thead>
<tr>
<th>Current use(s)</th>
<th>Post-(university)entry screening to identify university students in need of English support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test format/question types</td>
<td>C-test (text completion), Cloze elide (speed reading). A human-scored writing (argument essay) component can be added if required. See ** for validity evidence</td>
</tr>
<tr>
<td>Difficulty level</td>
<td>IELTS 5.5-8 (approx.)</td>
</tr>
<tr>
<td>Reliability</td>
<td>Internal consistency (Cronbach’s alpha): C-test: .93, Cloze elide: .96*</td>
</tr>
<tr>
<td>Parallel forms</td>
<td>6 statistically equated versions</td>
</tr>
<tr>
<td>Link to other measures</td>
<td>Formally linked to Diagnostic English Language Assessment (DELA) which in turn can be equated to IELTS levels if required.</td>
</tr>
<tr>
<td>Score report</td>
<td>3 classifications: Proficient, Borderline, At risk plus optional profile of writing ability</td>
</tr>
<tr>
<td>Time required</td>
<td>30 minutes or one hour if Writing is included</td>
</tr>
<tr>
<td>Cost</td>
<td>Site licence for groups $8,000 or per capita fee of circa Aus$4.00 (based on estimated candidature of 2,000) plus $22.00 for writing (if used)</td>
</tr>
<tr>
<td>Delivery</td>
<td>Online (either in supervised environment or self-access mode)</td>
</tr>
<tr>
<td>Equity</td>
<td>Sample tasks provided on the test, plus candidate handbook</td>
</tr>
</tbody>
</table>


The Academic English Screening Test (AEST) or Post-entry Assessment of Academic Language (PAAL), was designed to screen the language proficiency of university students from diverse language backgrounds including both native and non-native speakers of English, following admission to the university. The AEST serves to flag those who may experience such language difficulties and alert to them to the need to attend to their English language development. The test is pitched at intermediate to very advanced learners.

Validity

The test is made up of three sections: a text completion task, a speed reading task and an optional academic writing task. The 20-minute text completion task is a C-test, a format developed by Klein-Braley (1985) in which every second word in a series of short written texts on a variety of topics is partially deleted and students reconstruct the texts by filling the gaps. The 10-minute speed-reading task requires students to read a text of approximately 1000 words and identify superfluous words that have been randomly inserted (see Elder & von Randow, 2008). The optional writing task, administered online for score verification and diagnostic purposes, is a 30-minute argumentative
essay (see Elder, Knoch, & Zhang, 2009 for further detail). The first two components are automatically scored whereas the writing task, if used, is scored by trained raters using a three-category analytic rating scale.

Although the first two components are indirect measures, studies investigating the cognitive processes of test takers completing such tasks (e.g. Matsumura, 2009) show that test takers draw on a wide range of linguistic knowledge to complete them, including lexical, grammatical, lexico-grammatical, syntactic and textual knowledge. Statistical evidence for the validity of AEST as a language proficiency measure comes in the form of strong correlations (from 0.7 to 0.8) between scores achieved by the trial population on these tasks and on another professionally validated measure of academic English, the Diagnostic English Language Assessment (DELA) which is itself a proven predictor of future academic performance (Elder & Knoch, 2009; Knoch et al. in press). A limitation of the test is the lack of any component directly testing listening or speaking ability. Although high correlations with DELA listening comprehension component were found, there is no evidence indicating the predictive power of the AEST/PAAL in relation to speaking.

Reliability
Internal consistency measures for the C-test and cloze elide tasks are high at 0.93 and 0.96 respectively. Six statistically equated parallel forms of these tasks are available (see Elder & Knoch, 2009 for details of the equating process). Inter-rater reliability of the writing task is acceptable but this will always depend on the quality of raters and the training process which can be conducted either in-house by the relevant institution or out-sourced to the LTRC.

Comparability
AEST scores have not to date been directly linked to those of other widely recognized language proficiency measures such as IELTS, but there are established links to DELA at the University of Melbourne which has been equated to IELTS and can be therefore used to draw inferences about equivalent IELTS levels.

Utility
The test is quick to complete taking only one hour with the writing task included and 30 minutes if not. Feedback on the C-test and cloze components is immediate and is reported in the form of the 3-way classification: Proficient (language proficiency is sufficient for academic study), Borderline (in need of further language development) or At risk (of academic failure) and a link to local institutional guidance regarding avenues for further language support. Students can be classified differently depending on what the host institution deems (in the context of a standard setting exercise) to be an appropriate language level or levels for its particular purpose. What advice is given or action is taken with respect to the various thresholds on the test is a matter for institutional policy.

A diagnostic profile of writing strengths and weaknesses (see Appendix D for an example) can also be included on the report if the writing task is administered and marked as part of the procedure, but since the scoring is done by human raters rather than an automated scoring engine, this would delay the feedback by a number of days.

---

8 Such links could be formalized via a concordancing study involving the comparison of IELTS scores achieved by AEST/PAAL candidates.

9 The benchmark for setting standards at the University of Melbourne is based on performance on the DELA test, which was administered concurrently at the trial stage. The “at risk” cut-score was set at the point below which low scoring students on the AEST had the highest probability of also scoring at an unsatisfactory level on the DELA, on which standards had previously been set.
A single form of the AEST can be site-licenced to institutions at a cost of $8,000 (regardless of test taker numbers) but additional costs are involved if changes to the delivery platform are requested (e.g. the addition of the writing component). Such costings can be negotiated with the LTRC. Licensing fees could potentially be recouped from test takers on a per capita test registration basis.

The test format lends itself well to customization, and could be redesigned to reflect language ability for use in different domains, with the use of professionally oriented rather than academically oriented text-types for example. A systematic validation process would nevertheless be needed to establish the fitness of any new test form for its intended purposes.

Equity

A possible limitation of AEST/PAAL is its lack of face validity or obvious relevance to the target language use domain\(^{10}\). However a user handbook is readily available explain the basis for the test and providing examples of item types. Sample tasks are in also provided at the beginning of each section of the test.

There are no consequences for performing poorly on the test, which is simply a means of raising candidates’ consciousness regarding their level of language proficiency. However there may be negative consequences should the test be used for other more high stakes purposes, such as selection. One such consequence could be a tendency for potential candidates to focus on practising the tasks on the test rather than spending time on the kind of skills-based activities more likely to result in improved language proficiency for the relevant academic or professional purpose.

2. Aptis

<table>
<thead>
<tr>
<th>Table 2. Aptis (<a href="https://www.britishcouncil.org/exam/aptis">https://www.britishcouncil.org/exam/aptis</a>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact: +44 161 957 7755 or enquiry form on the website</td>
</tr>
<tr>
<td>Current use(s)</td>
</tr>
<tr>
<td>To evaluate general English skills for a range of purposes</td>
</tr>
<tr>
<td>Skill coverage/test components</td>
</tr>
<tr>
<td>Grammar, Vocabulary, Listening, Reading, Writing, Speaking, (can be customised in different combinations)</td>
</tr>
<tr>
<td>Test format/question types</td>
</tr>
<tr>
<td>Grammar: 25 multiple choice questions, Vocabulary: 25 questions</td>
</tr>
<tr>
<td>Listening: 25 questions, Reading: 4 tasks, Writing: 4 tasks, Speaking: 4 parts. Note that writing and speaking are scored online by human raters.</td>
</tr>
<tr>
<td>Validity evidence</td>
</tr>
<tr>
<td>See O’Sullivan (2015a)*; O’Sullivan (2015b)**; and other research papers on website</td>
</tr>
<tr>
<td>Difficulty level</td>
</tr>
<tr>
<td>CEFR A1 to C</td>
</tr>
<tr>
<td>Reliability</td>
</tr>
<tr>
<td>Average item consistency reliabilities from 0.82 to 0.91 across operational versions of listening reading and core (vocab and grammar) sections.****</td>
</tr>
<tr>
<td>Inter-rater reliabilities of 0.89 to 0.97.****</td>
</tr>
<tr>
<td>Parallel forms</td>
</tr>
<tr>
<td>Currently approximately 2,000 items in the test task and item bank; System designed to create a range of task and item types for each paper.</td>
</tr>
</tbody>
</table>

\(^{10}\) Note that face validity, or appearance of validity, is not to be confused with actual validity, since the perceptions of lay persons, while important, are an insufficient basis for determining whether a test actually measures what it purports to measure (Davies et al. 1999).
**Aptis (General)** is an English language proficiency test developed by British Council. The test assesses Grammar and Vocabulary, and four language skills, Listening, Reading, Writing, and Speaking. Institutions can choose any of the four language skills in different combinations. It is used by corporates for diagnostic or recruiting and promotion purposes, governments for evaluation purposes in general, or schools and universities for diagnostic purposes. Aptis for Teachers and Aptis for Teens are also available.

### Validity

The test is broad ranging in its coverage and structured to assess different language skills, listening, reading, writing, and speaking with grammar and vocabulary at its core. The theoretical basis of the test design is outlined in a report by O’Sullivan (2015a) and results of trial feedback are reported in O’Sullivan (2015b).

The Grammar section requires selection of the most appropriate word to complete a sentence or a short dialogue, and the Vocabulary section measures knowledge of synonyms, similar meanings, and collocations (e.g., bare fact, artificial light). The Listening component measures test takers’ understanding of monologues and dialogues of different lengths and on different topics. The Reading section is divided into 4 tasks each involving fixed choice options: choosing the most appropriate word to complete a short passage (e.g., a letter), ordering sentences to make a story, filling gaps in a text drawing from a provided list of words, and reading a passage and choosing a
paragraph heading from options provided. Writing comprises 4 parts and all are thematically linked

to a common topic. For example, candidates might be asked to fill in a form, write a letter, diary or

blog entry and email related to the theme of travel. Lastly, the Speaking section includes 4 parts:
speaking about oneself, describing a picture and answering the relevant questions, comparing two
pictures and answering two subsequent questions, and answering questions relevant to a picture
provided.

The test claims to assess learners across the A1- C range but does not discriminate within the upper
range, which may be critical for determine readiness for T&I work. The designers are currently
working on extending the range of tasks and items to encompass these upper levels, but this work is
not yet completed (Judith Fairbairn, personal communication).

Reliability
An item bank of 2,000 items has been developed. O’Sullivan (2015b) indicates that the test went
through 500 formal trials. Acceptably high internal consistency reliabilities for listening, reading and
core (grammar and vocab) components are reported in O’Sullivan and Dunlea (2015). Intra-rater
reliabilities for the human scored speaking and writing components are also acceptably high. (Raters
are trained online and the reliability of their scoring is checked at each administration).

Comparability
A standard setting study linking the test’s reporting scales to the CEFR has been conducted (see
O’Sullivan, 2015c).

Utility
The test is computer-based and can be administered using a tablet or computer, or via telephone. It
requires 3 hours to complete if all four skills are tested. Immediate scoring is only possible for the
receptive components of the test.

Score reports (see example in Appendix D) include a scaled score from 0 to 50 for each language
component tested, together with the corresponding CEFR skill descriptors.

A site licence is available for institutional purchases at a cost to be negotiated with the relevant
agency depending on the components of the testing package that are required.

Equity
Careful attention was paid to test taker feedback in refining aspects of computer delivery that
caused problems for some candidates (e.g. font size of the reading texts).

Videos with introductions for each section and demonstration tests for all sections are available on
the website.

Candidates are allowed a limited amount of pre-performance preparation time for both writing and
speaking (the time is built into the response times). Note that the time allowed for responding to all
items and tasks is carefully controlled to ensure a similar test experience for all candidates.
The CaMLA English Placement Tests assesses general receptive language proficiency in listening comprehension, grammatical knowledge, vocabulary range and reading comprehension. It is primarily used to place ESL students into appropriate ability levels and classes. It can also be used to determine suitable work assignments in occupational settings where English proficiency is required.

Validity

The test’s skills coverage is Listening, Grammar, Vocabulary, and Reading, and the items are situated in a variety of domains: educational, social, occupational, and personal. In Listening, there are two types of items: selecting the most appropriate response to a question or a statement and answering a question about a short conversation between two speakers. Grammar measures test takers’ grammatical knowledge in using the correct form of a word or a phrase in given context. Similarly, Vocabulary measures test takers’ vocabulary range; ability to select the correct word that correctly completes a sentence given. In Reading, sentence-level reading comprehension and comprehension of two reading passages of different lengths are assessed. All questions on the test are multiple choice with three options for the listening section and four options for all other sections. The productive skills, speaking and writing, are not assessed by this test.

A development report by Walter and Hentschel (2013) outlines a revision process for the test which was originally developed in 1972 for in-house use at the University of Michigan and then made available for use at other institutions in 1987. The report describes the subskills measured by each test and outlines the process of compiling the pilot test forms and pilot testing results. Although the
items properties show strong discrimination between different levels of ability, no concurrent or predictive validity evidence is provided to buttress these findings.

A factor analytic study (Walter and Hentschel, 2014) exploring the relationship between different sections of the test found that the test data did not show a clear separation between the different skill components. This resulted in a decision to report a composite score on a single scale representing receptive language ability rather than separate scores for reading and listening.

Reliability
Evidence of acceptable test reliability for items in the pilot test is found in Walter & Hentschel (2013). There are 6 parallel forms available which are set at the same level of difficulty in the paper-based test, and the computer-based test contains the same content as the paper-based one. A study (CaMLA, 2014) on comparing the paper-based and computer-based tests using two parallel forms shows that the two different delivery methods work uniformly, and the participants’ scores were strongly correlated (.90). Each form has unique content without shared items across the forms. Evidence for test reliability is provided in the report by Walter and Hentschel (2013).

Comparability
A study on the linking of CaMLA English Placement to the CEFR is available (see CaMLA, n.d.). The study shows comparison of the CaMLA test cut scores with the CEFR at a range of levels from A1 to C1. The alignment process involved eliciting the judgements of experts (assessment managers and directors, and developers of CaMLA) on the relationship between the CEFR levels and the cut score judgments for the test.

Utility
The test is computer-based and requires only 60 minutes to complete.

The CaMLA EPT is sold to institutions, which purchase test credits and also receive an administrator’s license. The institution pays CaMLA for a testing package, and can then provide access to individuals or groups to take the test online. The purchasing institution is at liberty to charge people they authorize to take the test. The costs of the testing packages are reasonably economical at US$595 for 100; US$148.75 for 25 with a minimum purchase of 25. This amounts to approximately US$6.00 per capita.

Score reports are generated for institutional use only and include only the overall CaMLA score, rather than being broken down by skill (see rationale under Validity above). The report (see example in Appendix D) is visible to the institution immediately after a candidate has taken the test. The score report does not include the CEFR levels, but the comparison of the test’s cut scores to CEFR level scores is available on the CaMLA website.

Equity
Sample items and a 4-minute video of introducing the test and the sample items for each section are available on the website. Practice tests are not provided.

It is strongly recommended by the test designers that the test be proctored to preserve its security for the purchasing institution, so that scores remain meaningful over time. The scores should not be reported to external institutions; the test is intended for internal use by the purchasing institution.
Table 4. Diagnostic English Language Needs Assessment (DELNA) Screening

<table>
<thead>
<tr>
<th>Current use(s)</th>
<th>Post-entry screening to determine EAP needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill coverage/test components</td>
<td>Reading, Vocabulary</td>
</tr>
<tr>
<td>Test format/question types</td>
<td>Cloze elide (speed reading), Vocabulary (matching words)</td>
</tr>
<tr>
<td>Validity evidence</td>
<td>Elder &amp; Erlam (2001)*; DELNA Screening test (speed reading and vocabulary) correlated well with DELNA Diagnosis (Listening, Reading, Long Academic Writing) test and a short academic writing task: .82; DELNA Diagnosis results to Grade Point Average (GPA)</td>
</tr>
<tr>
<td>Difficulty level</td>
<td>Approx IELTS 5.5 – 8</td>
</tr>
<tr>
<td>Reliability</td>
<td>See Elder &amp; Erlam (2001)</td>
</tr>
<tr>
<td>Parallel forms</td>
<td>Several parallel forms available</td>
</tr>
<tr>
<td>Link to other measures</td>
<td>No formal links to other recognized English measures (apart from DELA)</td>
</tr>
<tr>
<td>Score report</td>
<td>Proficiency, borderline, at risk classifications only</td>
</tr>
<tr>
<td>Time required</td>
<td>30 minutes (Time reading: 10 minutes, Vocabulary: 7 minutes)</td>
</tr>
<tr>
<td>Cost</td>
<td>Free for Auckland University students; Site licence at approx.: $8.00 per student.</td>
</tr>
<tr>
<td>Delivery</td>
<td>Online. Customisable delivery platform negotiable.</td>
</tr>
<tr>
<td>Equity</td>
<td>Information and test-taker handbook on DELNA website. <a href="http://www.DELNA.ac.nz.co">www.DELNA.ac.nz.co</a></td>
</tr>
</tbody>
</table>


The Diagnostic English Language Needs Assessment Screening Test was a precursor of the AEST/PAAL, developed as a preliminary screen for both native and non-native English speaking undergraduates and doctoral students following their entry to the university to determine who could be exempted from a more time consuming and costly academic English test, including both listening, reading components developed as part of DELA at the University of Melbourne and a locally developed writing task.

Validity

The screening test is pitched at a range of proficiencies from advanced to native-like (in the approximate range of 5 to 8 on IELTS). It includes a vocabulary and reading (speeded) component. The vocabulary component uses multiple-choice format to test understanding of isolated words selected from the academic word list. The speed-reading test uses a form of the cloze elide/speed reading format, that was later adapted for the AEST (see above), involving deletions of superfluous words inserted in the text. The coverage of the vocabulary test is considerably narrower than that of the C-test word completion format (used on the AEST), which draws not only on word knowledge but also textual, syntactic and pragmatic competence. Nevertheless the two components combined have, like the AEST, been demonstrated to correlate strongly with overall and component scores on the more time consuming DELA test of academic English.

The procedures followed to establish cut-score on the screening test mean that threshold level on the test is likely to accurately distinguish between linguistically at risk students and those who are unlikely to encounter problems with academic English.
Reliability
All items on the test meet acceptable standards of discrimination and overall reliability is high. Six statistically parallel forms of the screening tasks have been developed.

Comparability
Apart from the links established with the associated diagnostic procedure DELNA, there are no formal links to other standardized tests.

Utility
A limitation of the test is that, while it is highly efficient timewise (30 minutes) and can offer immediate feedback, it does not measure speaking or listening skills and the other two macroskills are assessed indirectly, rather than via communicative tasks. It thus gives only a broad indication of language level. If broader and deeper coverage and more detailed feedback is desired it would be advisable to administer it in combination with the optional writing component and an additional measure of speaking proficiency (where relevant).

Like DELA, score reports currently classify learners at three different ability levels, Proficient, Borderline and At risk and while the cut-scores between levels could be reset for different purposes, the screening test on its own does not give diagnostic information.

The test can be site licensed to institutions for a fee of $8.00 per student plus an additional set up fee of $550.00. A group discount option is also available on a sliding scale starting at $8,500 for up to 1,500 students. Additional (human-scored) writing components can be added for additional amounts depending on whether a local rater training package and training support is included. The web-based platform can potentially be customised in line with institutional needs, again for an additional fee.

Equity
The DELNA website provides ample information about the test and its purpose as well as handbook with samples of test items.

5. DIALANG

Table 5. DIALANG (http://www.lancaster.ac.uk/researchenterprise/dialang/about)
Contact: g.smith@lancaster.ac.uk

<table>
<thead>
<tr>
<th>Current use(s)</th>
<th>Broad skill coverage. Separate sub-tests for Listening, Reading, Writing (indirectly), Grammar, Vocabulary.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill coverage/test components</td>
<td>30 items per sub-test. 18 item types: including multiple choice, drag and drop matching, re-organisation, highlighting/underlining, insertion, deletion, thematic grouping, mapping and flow charting</td>
</tr>
<tr>
<td>Test format/question types</td>
<td>See Alderson (2005)*</td>
</tr>
<tr>
<td>Validity evidence</td>
<td>Wide-ranging (CEFR A1-C2). 3 levels of each sub-test (easy, intermediate and difficult)</td>
</tr>
<tr>
<td>Difficulty level</td>
<td>High. See Alderson (2005)*</td>
</tr>
<tr>
<td>Reliability</td>
<td>Not available.</td>
</tr>
<tr>
<td>Parallel forms</td>
<td>CEFR</td>
</tr>
<tr>
<td>Link to other measures</td>
<td>Separate report for each sub-test. Provides CEFR level with ‘Can do’ descriptor, opportunity to review answers by sub-skill, comparison between actual and self-assessment and advice on learning strategies.</td>
</tr>
<tr>
<td>Score report</td>
<td>Flexible. Level finder tests: approx. 15 minutes. Other test components:</td>
</tr>
<tr>
<td>Time required</td>
<td></td>
</tr>
</tbody>
</table>


The Dialang test is, as its name suggests, a diagnostic test of language ability which provides information about learner strengths and weaknesses. It is available in English and 14 other languages. The English version is what we are concerned with here.

**Validity**

The test is in fact a battery of 5 separate components, each containing 30 items testing reading, listening, writing and the foundational skills of grammar and vocabulary respectively. The theoretical rational for each test component is explained in separate chapters of Alderson (2005) and includes a detailed account of the discourse forms (for reading listening and writing) and sub-skills tested. Writing is assessed indirectly, via gap fill tasks where the candidate chooses the correct response from options provided or types in the correct word or phrase to fill the gap. The focus of the writing component is on mechanics (accuracy of spelling and grammar), appropriate textual conventions, the ability to recognize appropriate text conventions or distinguish between formal versus informal registers and to detect coherence and cohesion markers. While all the areas of competence tested on Dialang are relevant for using English in T&I contexts, there is no opportunity for extended text production by the candidate. There is thus a risk that the measure may either overestimate or underestimate the candidate’s ability. The test also lacks a speaking component. Evidence for relationship between various sub-tests is presented in Alderson (2005). No criterion related validity evidence is offered other than the links to the CEFR descriptors.

Dialang caters for a full range of ability levels. There is an easy, intermediate and difficult version of each test. The most suitable sub-test for the candidate is chosen on the basis of two preliminary placement or ‘level-finder’ tasks – one of which is a self-assessment whereby the learner answers Yes or No to a series of Can Do statements, about the chosen skills, for example: ‘I can understand short simple personal letters’, which are associated with levels of the CEFR. The second task is to complete a vocabulary size test requiring the candidate to distinguish real words in English from a list of both real and pseudo words. The level of performance on these two tasks combined is used to determine which test level is best suited to the learner’s ability. Both tasks are reported to be sufficiently predictive of learner proficiency (Huhta et al. 2002 p. 40) to make it likely that the test presented to the candidate will be at the appropriate level of difficulty.

**Reliability**

Reliability evidence for each subskill is provided in Alderson (2005). Following the piloting process and removal of poorly functioning items, all subcomponents of the test performance showed acceptable levels of reliability. No parallel forms of the test are however available, meaning that unless a candidate moves to another level before resitting the test, s/he may receive the same items more than once, with the risk of a practice effect artificially boosting test scores in the event that the two sittings are close enough for the candidates to remember previous answers.

**Comparability**

A chapter in Alderson (2005) offers a clear account of the mechanisms for establishing links between the test items and the various CEFR levels, using two different methods to arrive at cut-scores.
between levels. Results showed that empirical item difficulties corresponded well with the judged CEFR levels of items and hence that the levels assigned to test scores are defensible.

**Utility**
The test is free and can be accessed directly by individuals logging into the website https://dialangweb.lancaster.ac.uk/setals. The website is easy to navigate. The time commitment required for all test components is substantial but they can be completed at different times at the candidate’s convenience. The preliminary level finder tests take 10-15 minutes to complete and each subcomponent requires a further 30-40 minutes at the candidate’s own pace. Candidates can opt to receive immediate feedback on performance on an item-by-item basis as they do the test.

Separate, detailed score reports are produced for each sub-test. Each report indicates the level achieved on the CEFR and provides a description of what typical learners can do at that level and the opportunity to review item responses organized by sub-skills (e.g., for Reading, how many inferencing, reading for detail or for main idea items were answered correctly). A comparison between self-assessed and actual performance is offered as well as advice on strategies for improving the particular skill. Each component of the feedback can be accessed at will by the candidate by selecting from a series of buttons. Feedback gathered during the trialling process (see Alderson, 2005) indicated that the advice function of the score report was perceived to be very worthwhile, but the value of self-assessment feedback was not widely appreciated. The problematic nature of this kind of feedback was confirmed by Chapelle (2006).

**Equity**
Test Instructions are offered in many different languages and the candidate can choose the preferred language making it less likely that misunderstanding of instructions will interfere with a candidate’s ability to complete the test items as intended. The test delivery system is highly flexible, meaning that the candidates can do the test at their own pace and choose what skill to test and when. While this makes absolute sense for a low stakes self-assessment, the test designers caution that the test should not be used for other more high stakes certification purposes as the testing conditions are not standardized.

**6. Education First Standard English Test (Plus)**

<table>
<thead>
<tr>
<th>Current use(s)</th>
<th>To assess learners’ proficiency levels for academic and professional purposes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill coverage/test components</td>
<td>Listening, Reading</td>
</tr>
<tr>
<td>Test format/question types</td>
<td>Listening: Monologue and dialogue with multiple choice and matching items Reading: Single passages and two paired short passages with multiple choice questions</td>
</tr>
<tr>
<td>Validity evidence</td>
<td>Correlations with TOEFL iBT: Reading (.85) and Listening (.85); with IELTS: Reading (.91) and Listening (.90)<strong>/</strong>*</td>
</tr>
<tr>
<td>Difficulty level</td>
<td>Beginner to Advanced; CEFR A1 to C2</td>
</tr>
<tr>
<td>Reliability</td>
<td>Overall section reliability: EFSET Plus Reading (.90 to 95) and Listening (.88 to 94) over two administrations</td>
</tr>
<tr>
<td>Parallel forms</td>
<td>A computer adaptive test; items presented to test takers are drawn from a large item pool</td>
</tr>
<tr>
<td>Link to other measures</td>
<td>CEFR*</td>
</tr>
<tr>
<td>Score report</td>
<td>Reported on the EF scale (1 to 100) with corresponding CEFR levels</td>
</tr>
<tr>
<td>Time required</td>
<td>EF Standard English Test: 50 minutes; EF Standard English Test Plus: 2 hours</td>
</tr>
</tbody>
</table>
Education First (EF) is an international education company specialising in language training, educational travel, academic degree programs, and cultural exchange with its schools in 45 cities worldwide. The EF Standard English Test (SET) was initially developed as a placement, advancement, or certification test for internal use within the EF, but is now publicly accessible. The test is designed to measure English learners’ proficiency levels for academic and professional purposes. The test reports scores only to individual test takers; therefore, it can be used as a practice before taking other standardised tests such as the Cambridge English FCE, TOEFL, or IELTS as stated on the website. There are three types of EFSETs, EFSET Express, EFSET, and EFSET Plus, which are 15 minute, 50 minute, and 120 minute tests respectively. The different types are distinguished by the number of texts and items included, with the longer tests (EFSET and EFSET Plus providing more stable ability estimates for high stakes purposes. Speaking and Writing tests are currently in design phase.

Validity
There are three types of items for Listing: Monologue with multiple choice or multiple option questions, dialogue with multiple choice or multiple option questions, and matching, and two types of items for Reading: Single passage with multiple choice or multiple option questions with completions and two paired passages with 6 fixed format questions.

Details of the test development process, trials, and foundations of theoretical model underpinning the test design are outlined in a technical report by Education First (Education First, 2014). The actual test development is carried out using the assessment engineering approach (Luecht, 2012).

Two correlation study reports (Luecht, 2015a, 2015b) with IELTS and TOEFL showed very strong correlations (.91 for IELTS Reading, .90 for IELTS Listening) and strong correlations (.85 for both TOEFL iBT Reading and Listening) respectively.

Reliability
The test is a computer adaptive test with items drawn from a large pool. Reliability estimates calculated for two test administrations were high at .90 and .95 for Reading, and .88 and .94 for Listening.

Comparability
Scores are formally aligned to CEFR levels. This standard setting and mapping to the CEFR level process involved 15 language subject matter experts’ judgements.

Utility
EFSET is free to users and administered as a web-based tool and works with all recent browsers except for Internet Explorer 9 or below. However researchers trying out the test encountered
persistent technical difficulties, including freezing of the computer screen. These difficulties were not resolved even after technical assistance was sought.

Scores reports include an EF scaled score with a general level (e.g., upper intermediate) in conjunction with the corresponding CEFR levels for each skill (e.g., B1 reading, B2 listening), which are linked to their descriptions on the web.

**Equity**

No sample tests are available on the website, but test takers can freely take the test repeatedly at no cost. A practice effect is unlikely given that randomized selection of items by the computer at each administration.

### 7. Oxford Online Placement Test

**Table 7. Oxford Online Placement Test**  
Contact: +44 1865 354500 or customerservice.eltonline@oup.com

- **Current use(s)**: Mainly to place learners in appropriate classes.
- **Skill coverage/test components**
  - 2 components: *Use of English* (vocab and grammar in context) and *Listening*
- **Test format/question types**
  - *Use of English* (30 items): Completing a conversation by choosing a word or phrase, matching phrases with similar meaning, gapfill; *Listening* (15 items): Multiple-choice responses to audiorecorded dialogues.
- **Validity evidence**
  - Information about theoretical foundations of test construct available online (see Purpura, n.d.* on website): e.g., relationship between form, semantic meaning and pragmatic meaning.
- **Difficulty level**: CEFR A1 to C2
- **Reliability**
  - Test-retest results: .91 (total score) .87 *Use of English*; .82 *Listening* (see Pollitt**).
- **Parallel forms**: N/A. Test is computer-adaptive and draws on large item bank (see Pollitt**).
- **Link to other measures**: Scores aligned to the CEFR via conceptual and empirical mapping process*
- **Score report**: Score (0-120) linked to CEFR level plus descriptive profile
- **Time required**: 45 - 90 (default) minutes
- **Cost**: £4.00/unit
- **Delivery**: Via internet link accessible on site with test-taker login details or remotely by means of email
- **Equity**: Online practice tests freely available.


The Oxford Online Placement Test (OOPT) is used primarily to aid decisions about student placement in English classes but it could also be used as a general gauge of an individual’s ability.

**Validity**

The computer adaptive design of the test relies on a large bank of test items that are matched (through an extensive trialling process conducted in schools around the world) to a broad range of ability levels, for any level on the Common European Framework of Reference (CEFR) from beginner to very advanced.
The test’s skills coverage is limited to two areas of language ability: *Use of English* and *Listening*. The *Use of English* section is designed to knowledge of grammatical forms and the literal and implied meanings (at the word, phrase and sentence level) that these forms encode in particular contexts of use. The listening section measures understanding of literal and implied meanings in the context of both dialogues and monologues of different lengths and on different topics. An interesting feature of the listening test (which is quite unusual in language testing contexts) is that British and US accents are used for the recordings and candidates can select either one or the other or a mix of the two. Australian accents are not represented. A paper by Purpura (see table above) outlines the theoretical rationale for the test design emphasizing its focus on contextualized meanings, which were emphasized in the proficiency-related KSAs for the T&I profession summarized in Section 1 of this report.

Speaking, reading and writing skills are not directly measured by this test. Although the grammar and vocabulary skills tested in the *Structure and Language Use* section have been found to be good predictors of reading and writing skills and there may well be a relationship between listening skills and of speaking ability, no empirical evidence is provided in this regard.

**Reliability**

Evidence for test reliability is provided in Pollitt (see reference in above table) and shows that the test produces acceptably reliable estimates of language ability although it is stressed that scores at the borderline between one band level and another need to interpreted with caution.

Alternate forms of the test are not needed as the test is computer adaptive and draws on clusters of pre-tested items at each difficulty level that are presented randomly to test takers. IELTS equivalences can be calculated with reference to a score comparison chart (see page 6 above).

**Comparability**

Scores on the test have been formally aligned to Common European Framework of Reference, a widely used scale providing descriptions of language ability through a range of levels from A1 (Basic) to C2 (Proficient user). The alignment process involved eliciting teacher judgements of student ability against the CEFR level descriptors and mapping these judgements to the same students’ scores on the placement test (see Pollitt, n.d., referenced above, for further details).

**Utility**

The test draws on a bank of over 1500 questions and uses computer-adaptive techniques to tailor the level of test questions to an individual student’s ability thereby reducing the time required (typically 45-60 minutes minutes) to make an accurate estimate of the student’s level.

Score reports include two separate sub-section scores to allow users to identify areas of relative strength or weakness for planning purposes and a total (average) score that is linked to a CEFR level and associated description (see Appendix D).

The test is sold via the Oxfordenglishtesting.com Online shop. It can only be purchased and administered by organization accounts and cannot be used by self-study students. Single-use licences (i.e. 1 student for 1 test) are sold for a relatively low per capita fee of £4.00 which could be recouped from NAATI candidates. There is a minimum purchase of 10 tests for the first order, with a discounted price offered to institutions depending on how many tests are purchased at once. The tests have no expiry date which would allow for advance bulk purchase of tests for an entire annual NAATI candidature regardless of when candidates are taking the test.
Equity

Practice tests are provided on the Oxford OOPT website along with other information about the test which is expressed in a manner that test takers are likely to find easy to follow. A free trial is available for purchasing institutions. There is some latitude for candidates in how long they take to complete each section and listening items can be played more than once. While this may affect performance, it should be noted that the time taken by each test taker is included on the score report and could potentially be taken into account in the interpretation of test scores.

There appears to be no provision for individual self-access or any possibility for resitting the test should a candidate not reach the desired level.

8. Password

Table 8. Password

<table>
<thead>
<tr>
<th>Current use(s)</th>
<th>Admission to universities, placement in EAP courses, study planning, and post enrolment language assessment. Selection for professional purposes.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill coverage/test components</td>
<td>Language Knowledge with optional add-on skill components (Reading, Listening, Writing and Speaking).</td>
</tr>
<tr>
<td>Test format/question types</td>
<td>Knowledge test has 5 components with multiple-choice questions on different aspects of grammar and vocabulary knowledge tested mainly in sentence level contexts.</td>
</tr>
<tr>
<td>Difficulty level</td>
<td>CEFR A2 to C1 (lower end); IELTS 3.5 to 6.5</td>
</tr>
<tr>
<td>Reliability</td>
<td>.916</td>
</tr>
<tr>
<td>Parallel forms</td>
<td>Parallel versions created via randomized selection from a large item bank.</td>
</tr>
<tr>
<td>Link to other measures</td>
<td>Broadly predicts scores on IELTS and outcomes on other tests linked to CEFR (see ** below this table for evidence of such links)</td>
</tr>
<tr>
<td>Score report</td>
<td>Separate components and overall test scores are linked to CEFR levels statements</td>
</tr>
<tr>
<td>Time required</td>
<td>Knowledge test (Vocab and grammar) 45-60 mins. Reading and Writing: 1 hour and 45 minutes; Listening and Speaking: 1 hour and 15 minutes (Writing and Speaking are marked by human raters)</td>
</tr>
<tr>
<td>Cost</td>
<td>Aus$20-$45.00 per capita depending on the number of test modules taken.</td>
</tr>
<tr>
<td>Delivery</td>
<td>Online</td>
</tr>
<tr>
<td>Equity</td>
<td>Practice tests available for individuals and can be purchased online.</td>
</tr>
</tbody>
</table>


The Password test is designed to provide an efficient and accurate indication of adult learners’ English proficiency level to for language planning or placement purposes to determine whether they are at the level needed to embark on a university course delivered in English (or for a number of other possible purposes as listed in the table above). The Knowledge (grammar and vocabulary test)
consists of 60 items divided into five different parts covering different kinds of vocabulary and grammar knowledge tested via multiple-choice, true/false and sentence-matching item types. A professional (rather than academic) version of the test which offers a slightly different contextual frame for the same test items is also available. There are optional computer-scored Listening and Reading components as well as Writing and Speaking components delivered online and scored by human raters. The Knowledge test can be used in combination with one or more of these skill-based components or as a preliminary screen to determine whether a candidate is ready to take the skill-based components or to take other tests such, as the IELTS, designed for particular academic or professional purposes.

Validity
Password was developed on the basis of Weir’s (2005) socio-cognitive framework for test validation. Green (2011) justifies the use of a knowledge, rather than skills-based, test for screening purposes on the grounds that vocabulary and grammar are the building blocks of proficiency and broadly predictive of results on other tests. Evidence of such predictions can be found in ‘Password and the CEFR’ on the test website. A limitation of the knowledge test is that it measures knowledge mainly in the context of sentences rather than longer texts and thus places little emphasis on textual features such as coherence and cohesion. This may limit its predictive power in relation to the advanced levels of proficiency in the four skills that NAATI candidates may require. It seems unlikely that this would be resolved by a differently worded “professional version” of the test because the context provided for the items is similarly limited.

The test spans a range of proficiency levels from IELTS 3.5-6.5 (CEFR A2 – borderline C1) and is formally linked to the CEFR via procedures described meticulously on the relevant website. However while the test discriminates well at the lower end of the ability range it does not differentiate between learners along the C1 - C2 (proficient user) levels of the scale (or at levels above IELTS 6.5)11. Any advanced user of English (whether native or non-native speaker) would receive a Password Plus score at the top end of the test scale and informed of readiness to take the IELTS for university entry but no indication of any gradations of ability within this high score range. This is a serious limitation in situations such as interpreting and translating where the minimum language requirement is likely to be set higher than the minimum required level for embarking on university study.

Reliability
The pretesting of items in large-scale trials shows high levels of inter-item consistency at around 0.9. All items that do not meet acceptable parameters of difficulty and discrimination are discarded at the test development stage.

Multiple parallel forms of the test are generated by random selection of items spanning the range of difficulty levels from a large item bank. The bank is continuously supplemented with new items, a small number of which are included in each test version for piloting purposes. The existence of this renewable item bank ensures test security in that each individual is likely to receive a different set of items.

Utility
A single form of the test called Knowledge Test Solo can be used un-invigilated for self-screening by individuals for the sum of £10.00 GBP.

11 Whether this is necessary for NAATI screening purposes is a matter for professional determination.
Alternatively, the Password company can set up a partnering relationship with the relevant agency and provide them with a co-branded version of their tests which would be delivered in a proctored environment following their suggested procedures. Such an arrangement might include the receptive skills of listening and reading only (at around Aus$25.00 per capita), or all four skills (at $45.00). Alternatively the agency might opt for the Knowledge test only at $23.00 or the Knowledge test with Writing which could be scored either off-shore by Password trained raters or in-house. The fee for scoring the Writing test by the Password team is around Aus$6.00 per test. Fees for groups of test-takers are organized on a sliding scale whereby a large numbers of test-takers would result in slightly lower per capita fees.

Score reports for all test components indicate the level on a 9-point scale from Pre-Password to Password Plus that a candidate has attained. These levels are equated to CEFR levels from A1 to C1. There also advice about what action to take in relation to the score obtained (see Appendix D for an example for the Knowledge test).

**Equity**

Information is provided to test takers in English on the Passport website. Although instructions are in English only the item types on the all test components follow fairly familiar formats and should not present problems for test takers.

There is a potential fairness issue associated with the use of the un-invigilated version of the Knowledge test in the sense that candidates can “cheat” if they perceive the stakes to be high enough by consulting a dictionary or getting someone else to take the test in their place. There is also a risk that anyone who resits the test may improve due to previous practice with the same test items rather than due to any growth in language proficiency.

**9. Progress**


*Contact: +61 2 9454 2222 (Australian office for university, or personal and professional titles)*

<table>
<thead>
<tr>
<th>Current use(s)</th>
<th>To assess achievement over the course of an English instructional program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill coverage/test components</td>
<td>Skills assessed: Listening, Reading, Writing, Speaking, Grammar, Vocabulary</td>
</tr>
<tr>
<td>Test format/question types</td>
<td>60-65 items across 17 item types: Fill in the table, choose the right word/phrase, complete the dialogue, drag and drop, error correction, choose the right picture, short answer, listen to the conversation, describe image, short essay, read and then speak, listen and then write, listen and then speak, read and then write, passage comprehension</td>
</tr>
<tr>
<td>Validity evidence</td>
<td>High correlation between automated and human scores (.94*) and also with performance on Versant, TOEFL iBT, TOEIC and IELTS</td>
</tr>
<tr>
<td>Difficulty level</td>
<td>CEFR A1 to C2</td>
</tr>
<tr>
<td>Reliability</td>
<td>0.91</td>
</tr>
<tr>
<td>Parallel forms</td>
<td>Part 1 is computer adaptive so no parallel forms needed. Multiple equated forms of Part 2 test available</td>
</tr>
<tr>
<td>Link to other measures</td>
<td>CEFR and Global Scale of English (GSE)*</td>
</tr>
<tr>
<td>Score report</td>
<td>3 scores on Global Scale of English for Start, Middle and End of course</td>
</tr>
<tr>
<td>Time required</td>
<td>40-55 minutes</td>
</tr>
</tbody>
</table>
Progress is a four skills test of English language proficiency designed to support learners in improving their English by measuring their progress on the Global Scale of English (GSE). The test is divided into two stages, the first of which measures vocabulary, grammar and reading is computer-adaptive or, in other words, tailors the level of items presented to the candidate’s particular level of ability as determined by responses to the previous items. Via this procedure a rough estimate is made of the candidate’s overall level, so that the correct test for a more refined test of listening, speaking and writing as well as integrated skills items which combine these different skills. There are six test levels available for this second part of the test, which is not adaptive, but nevertheless involves automatically scored test responses.

Validity
The conceptual basis for the test is Assessment for Learning, an approach which sees testing as both informing teaching and being informed by it in a continuous feedback loop. The test is therefore, strictly speaking, designed to be used in the context of learning materials and formative assessment tasks along the lines of Pearson’s structured English syllabus which provides a detailed description of learning objectives for each level of the GSE scale. As well as the Pearson syllabus, The design draws extensively on the descriptive frameworks of the CEFR and encompasses both usage and use aspects of communicative competence or those relating to correct use of the linguistic system and those relating to effective communication (see Progress-Pearson, 2014). These aspects are assessed both separately and in combination.

Correlational analyses showed adequate separation between skill components, indicating that each measures different aspects of the test construct.

Although the test is aligned to a particular instructional pathway, scores of the trial population on the Progress test show high correlations with those on other recognized proficiency tests including Versant, TOEFL iBT, TOEIC and IELTS.

The difficulty range of the Progress tests is too limited for NAATI’s purposes as it is pitched to teachers of intermediate students preparing themselves for high stakes tests such as IELTS or TOEFL.

Reliability
The reliability of test items on each component is reported to be high in the range from 0.87 to 0.93. Human ratings on the various linguistic aspects were used to train an artificial intelligence engine in such a way as to produce ratings which correlate strongly (at 0.94) with human scores, a sign that the engine is ranking people as human judges would.

Utility
The test costs around $38.00 per unit and is administered either as a standalone to be used with any course or bundled with a specific Pearson course. The course leader controls how the test is taken and under what conditions, with different levels of supervision depending on students’ needs. The fact that the test is intended as an instructional tool limits its utility as a standalone measure of proficiency.
Score reports take the form of a skill profile with a separate score for each test component (listening, reading, speaking, writing, vocabulary and grammar) as achieved on tests taken at the start of a course, as well as midway through the course and at the end. The scores are reported on the Global Scale of English (GSE) (from 10-90) developed by Pearson to describe stages of learning achievement. The GSE serves a function similar to the CEFR and has been aligned to it. However the GSE the levels are more finely gradated and formulated giving teachers and learners a clearer sense of progress made. The relevant CEFR levels corresponding to the GSE scale are also provided along with their associated descriptors.

The skills profile on the score report can be used by teachers in conjunction with a “Can-Do dashboard”, a diagnostic tool which identifies particular learning objectives students could work on.

**Equity**

Bias analyses are conducted at the test design stage to remove ‘sensitive’ items that might be perceived as offensive or discriminatory by particular groups.

The test developers stress that Progress is a formative assessment instrument and not designed to be administered in a secure proctored environment. The score cannot therefore be used for certification purposes but instead serves simply to support and monitor students’ learning.

**10. Versant English Placement Test**

**10. Versant English Placement Test**

Contact: +1 450 470 3505 or sales@pearsonkt.com

- **Current use(s)**
  - To evaluate a student’s English communication skills for course placement or exit exam; to monitor student or employee progress and measure instructional outcomes; to benchmark language levels of employees, teachers and students for correct placement in training or remedial programs.

- **Skill coverage/test components**
  - Skills assessed: Listening, Reading, Writing and Speaking

- **Test format/question types**
  - 81 items: Read aloud, repeat, sentence builds, conversations, typing, sentence completion, dictation, passage reconstruction, summary & opinion.

- **Validity evidence**
  - 0.99 correlation between automatically and human scored samples.

- **Difficulty level**
  - CEFR A1 to C2

- **Reliability**
  - 0.96 Overall, 0.80-.95 other components. SEM 2.7*

- **Parallel forms**
  - 3 forms

- **Link to other measures**
  - CEFR (see the technical reports for Versant Speaking and Versant Writing)

- **Score report**
  - Overall and four diagnostic subscores (20 to 70) are reported together with corresponding levels of CEFR. Plus capability statements and study advice.

- **Time required**
  - ~50 minutes

- **Cost**
  - US$39.95/capita

- **Delivery**
  - Administered via computer after downloading the software available on the website. Score reports available within minutes of completion.

- **Equity**
  - Keyboarding skills assumed. Typing speed and accuracy are reported to aid interpretation of results as per Versant Writing and Speaking tests. Sample test available on website. Demo test can be requested via form on website. Validity of scores depends on verification of candidate identity at time of test administration.

The Versant English Placement Test is designed to measure test takers’ communication skills for course placement or as an exit exam. It is currently used for course placement by academic institutions, private language schools, and corporate training programs. The test uses selected components from Versant English Speaking and Versant Writing Tests (see below) but also contains unique tasks. The speaking and writing components act as a surrogate measure for listening and reading respectively as completing the tasks requires integration of the receptive skills.

Validity

A validation summary (see link under table above) reports on common features including shared specifications, tasks, rating criteria and scoring models between Versant Speaking and Writing tests. All three tests have the same theoretical underpinnings and hence can be said to measure similar constructs, albeit with different emphases.

The test has nine sections. There are 4 item types for the speaking part of the test: read aloud, repeat, sentence builds, and conversations. Listening prompts are recorded in a range of accents. The writing section has 5 item types: typing, sentence completion, dictation, passage reconstruction, summary & opinion. The item types are either identical or similar to those used in the Versant English Speaking, and the Versant Writing (see below). Test content is however more academic than in those tests, which are more professional in their orientation. A new open-ended summary and response task has been the subject of several validation studies and has been demonstrated to provide reliable and accurate automated estimates of ability.

A study comparing machine scored estimates with scores assigned by trained and reliable human raters found high correlations from 0.89-0.91 across the various scoring criteria indicating that the automated scoring system can accurately simulate human judgements on the relevant tasks.

The difficulty level of the test is not as high as that of the Versant English (Speaking) and Writing tests. Items on the Versant Placement test do not accurately measure performance above the B2 level. There is therefore likely to be a ceiling effect such that candidates the test does not provide useful information for those at the top of the ability range.

Reliability

The validation report indicates very high levels of internal consistency estimated with the split-half reliability method (see indices above). The Standard Error of Measurement is quite low (only 2.7 points on a 50-point score scale).

The test relies on randomized item selection for the creation of new forms, but the item bank is fairly small. There is only one fixed form of the test available for individual use.

Comparability

Scores are aligned to the Versant Speaking and Writing tests and, by implication, to the same CEFR level estimates. See the technical reports (Pearson Education, 2011, 2013) on the speaking and the writing for further information on the alignment process.
Utility
The test can be licensed to an institution or is directly available to individuals from a website at a fee of US$39.95 per capita. Licensed institutions can control the form in which scores are released to candidates or indeed if they release these scores at all. The test software can be downloaded from the website and installed on a computer. Headphones are needed and the test begins with a test to ensure that sound quality is adequate and the internet connection is stable enough to run the test. The test requires up to 50 minutes to complete.

Scores reports include overall and five subskill scores with descriptions of capabilities in addition to the corresponding CEFR level and descriptor (see example in Appendix D). Score reports are posted to www.VersantTEST.COM and can be accessed by test administrators and/or score users from a password-protected section of the website.

Equity
Since the test requires keyboarding skills a typing test is included at the start of the placement test which allows test takers to familiarise themselves with the keyboard and at the same time measures their typing speed and accuracy. If typing speed is below 12 words per minute and/or accuracy is below 90%, then it is likely that the candidate’s written English proficiency was improperly measured due to poor typing skills. The test administrator or individual test-taker may wish to take this into account when interpreting test scores.

A set of sample items for the speaking part is available, but not for the writing. A demo can be requested by filling in the form on the website.

Note that the test designers support the use of Versant scores for decision-making about learners only in circumstances where the identity of individuals taking the test can be verified at the time of administration. This means that self-assessments undertaken in an unproctored situation should be interpreted with caution.

11. Versant English - Speaking

Table 11. Versant English-Speaking (https://www.versanttest.com)
Contact: +1 450 470 3505 or sales@pearsonkt.com

<table>
<thead>
<tr>
<th>Current use(s)</th>
<th>To evaluate the spoken English skills for a range of purposes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill coverage/test components</td>
<td>Speaking (test components: Sentence mastery, Vocabulary, Fluency, and Pronunciation)</td>
</tr>
<tr>
<td>Test format/question types</td>
<td>63 items in six sections: 8 reading printed sentences, 16 repeating sentences verbatim, 24 short questions, 10 sentence builds, 3 story retelling, and 2 open questions</td>
</tr>
<tr>
<td>Validity evidence</td>
<td>Effective separation between NS and NNS groups; High correlations with TOEFL iBT overall (.64), CEFR (.84-.94), IELTS Speaking (.76); High correlation with human scores: .97*</td>
</tr>
<tr>
<td>Difficulty level</td>
<td>CEFR A1 to C2</td>
</tr>
<tr>
<td>Reliability</td>
<td>Test-retest results: .97*</td>
</tr>
<tr>
<td>Parallel forms</td>
<td>A set of 63 items are drawn at random from a large item pool.</td>
</tr>
<tr>
<td>Link to other measures</td>
<td>CEFR, TOEFL iBT Speaking, IELTS Speaking, and others*</td>
</tr>
<tr>
<td>Score report</td>
<td>Overall score and 4 diagnostic subscores (i.e., sentence mastery, vocabulary, fluency, and pronunciation) and skill descriptions are reported together with corresponding scores to CEFR, TOEFL Speaking, TOEFL total, and TOEIC.</td>
</tr>
</tbody>
</table>
The Versant English Speaking test is designed to measure how well a person can understand and speak spoken English on everyday topics and whether they can respond appropriately at a native-like conversational pace in English. It is currently used by academic institutions, corporations, and government agencies to evaluate their students’ or staff’s general skills in spoken English.

**Validity**

The test caters for a broad range of abilities, from beginners to advanced, and has 6 item types: reading, repeats, short answer questions, sentence builds, story retelling and open questions. The last item type, consisting of open questions, is not scored automatically but provides a sample for verification by human raters in the event of any automated scoring anomaly. The various test items provide multiple, independent measures that underlie facility with spoken English, including phonological fluency, sentence construction and comprehension, passive and active vocabulary use, listening skill, and pronunciation of rhythmic and segmental units. In the test description and validation summary report (Pearson Education, 2011), the theoretical foundations of test construct are outlined and it is stated that the test probes the psycholinguistic elements of spoken language performance rather than the social, rhetorical, and cognitive elements of communication. The reason for this focus, provided in the report, is to ensure that test performance relates most closely to the candidate’s facility with the language itself and is not confounded with other factors.

Validation studies of Versant Speaking show effective separation between NS and NNS groups and the differences between these groups were also used as the basis for devising scoring models.

A limitation of the test is that it measures monologic production in response to machine delivered stimuli rather than interactional ability. Tasks such as repeating or reordering sentences may seem rather artificial to test-takers. The test designers claim that the test is strongly predictive of such ability as evidenced in, for example, a moderately high correlation (0.78) between candidates’ overall Versant score and scores awarded to the same candidates on the face-to face IELTS speaking component.

**Reliability**

Each set of test items is drawn from a large item pool, and the report (Pearson Education, 2011) shows that the test yields reliable estimates (.97).

It should also be noted that correlation between human and machine scores was very high (at 0.97) supporting the notion that automated scoring can yield scores on a par with those assigned by human raters.

**Comparability**

Moderately strong correlations between Versant Speaking and other recognized English proficiency measures are reported including TOEFL overall, TOEFL iBT overall, TOEIC, TOEIC listening, Test of
Spoken English (TSE), TOEFL iBT speaking, and IELTS speaking (see Table above). These correlation indices were sought through parallel administrations. Versant scores have also been formally linked to CEFR levels, and this involved human raters judging responses to open questions in relation to the CEFR levels. Overall rater agreement with respect to these judgements was .89.

Utility
The test is automatically scored and scores are available online within minutes of completing the test. Score reports are very detailed and include both an overall score and subscores for each assessment component - Sentence Mastery, Vocabulary, Fluency and Pronunciation - on a scale from 20 – 80 points with a description of score meaning and the candidate capabilities that can be inferred from this score. The candidate also receives advice on what typical learners at the given level need to do to improve their English and an indication of how the Versant score lines up with CEFR levels, as well those on other recognized measures including the TOEFL and TOEIC. (No link is available to IELTS but such links can perhaps be inferred from concordance tables comparing scores across other major international tests)

The test can be administered by telephone or via computer (Mac or PC). When by telephone, a test taker is provided with instructions and an individual test form with the phone number to call and the test identification number. When via computer, the software can be downloaded from the website. The test begins after preliminary checks to ensure sound quality and internet stability are adequate. It takes 15 minutes to complete and costs US$35 per capita.

Pearson offers guidelines and materials (in the form of samples of speech elicited from test takers at different score levels) to aid decision-makers in determining cut-scores between satisfactory and unsatisfactory competence for the intended purpose. A fee is charged for this service.

Equity
A sample test paper and a sample call with performed responses are available on the website and a demo can also be requested by completing the form on the website; however, practice tests need to be purchased. One practice test form costs US$29.95 and three simplified quick practice tests can be purchased at US$19.95 each.

12. Versant English - Writing

Table 12. Versant Writing
Contact: +1 450 470 3505 or sales@pearsonkt.com

<table>
<thead>
<tr>
<th>Current use(s)</th>
<th>To measure how well a person can handle everyday and workplace English in written form (i.e., reading and writing skills)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill coverage/test components</td>
<td>Sentence comprehension and construction, passive and active vocabulary use, appropriateness and accuracy of writing assessed via integrated listening/reading and written response tasks.</td>
</tr>
<tr>
<td>Test format/question types</td>
<td>Test contains 43 items of multiple types: 1 Typing a passage, 20 sentence completion, 16 dictation, 4 passage reconstruction, 2 email writing. Scoring criteria: Grammar, Vocabulary, Voice &amp; Tone, Organisation, Reading Comprehension</td>
</tr>
<tr>
<td>Validity evidence</td>
<td>Effective separation between NS and NNS; High correlation with human scores: .98* Dimensionality analysis shows subscores measure different aspects of test construct</td>
</tr>
<tr>
<td>Difficulty level</td>
<td>CEFR A1 to C2</td>
</tr>
</tbody>
</table>
• Reliability: High reliability: 0.87-98 for subscores and 0.98 for overall score SEM 2.2*
• Parallel forms: Equivalent (in difficulty and content) versions of each item type are assembled from a large item pool for each test administration.
• Link to other measures: Correlation with TOEIC Listening (.68), TOEIC Reading (.70), TOEIC total (.72), CEFR*
• Score report: Score scale (20-80). Skill descriptions reported for Overall and 5 subcomponents, plus CEFR level, and advice for learners.
• Time required: 35 minutes
• Cost: U$29.95 per capita
• Delivery: Administered and automatically scored via computer after downloading the software available on the website. Purchase via site licence. Single form available for individual use.
• Equity: Keyboard skills assumed. Typing speed and accuracy measured to assist score interpretation. Sample items shown on website.


The Versant English Writing is designed to measure the test taker’s ability to understand spoken or written English and respond in writing appropriately on everyday and workplace topics. It is currently used for hiring or promotion, screening of teachers or teaching assistants, or progress monitoring of English writing training.

Validity

The theoretical foundations of the test construct are outlined in the test description and validation summary report (Pearson Education, 2013). Effective performance on the test relies on the psycholinguistic skills of automaticity and anticipation. See Pearson Education (2013) for how these skills are deployed in completing the various test item formats.

There are 5 item types: typing, sentence completion, dictation, passage reconstruction, and email writing. The business domain of the items developed includes announcements, business trips, complaints, customer service, fax/telephone/email, inventory, scheduling, and marketing/sales. Writing for academic purposes or writing in other specialist domains such as the law or medicine is not covered. Reading comprehension is measured to the extent that understanding of input material it is needed for writing responses, such as in dictation, summarizing a passage or answering an email.

The model used for scoring responses draws on the work of Grabe and Kaplan (1996). Three types of language knowledge proposed by these scholars are in line with the test’s subscore criteria of Grammar and Vocabulary (linguistic knowledge), Organisation (discourse knowledge), and Voice & Tone (sociolinguistic knowledge) respectively. Comprehension of spoken and written input is also assessed.

A technical report (Pearson Education, 2013) shows effective separation between NS and NNS groups as expected. Similar to the Versant English Speaking test, an analysis of the internal quality of the test involving a comparison between automated scores and human judgements was an integral part of the test validation showing a high level of alignment (i.e. an overall correlation of .98).

A dimensionality analysis showed that the different score components, while highly correlated, showed sufficient separation from one another to be regarded as representing different aspects of the test construct and provided justification for separate reporting of scores.
Correlations between Versant scores and those on other recognized English tests are acceptably high as shown in the Table above. This indicates that the tests are measuring similar skills.

Reliability
A study comparing the reliability of human scoring with that of automated scoring used on the operational versions of the test showed that the latter was more reliable than the former.

Each test administration draws on a large bank of items taking into consideration the item’s level of difficulty and its form and content in relation to other selected items. This means that different forms of the test are equivalent.

Comparability
Versant scores have been equated to those on the Test of English for International Communication (TOEIC) and have been formally aligned to CEFR level estimates. The former equivalences were established through parallel administrations of the two tests, and the latter involved eliciting experts’ judgements in mapping the selected writing samples to the A1 through C2 CEFR levels.

Utility
The software can be downloaded from the website and installed on a computer, whether Mac or PC. Headphones. The test requires up to 35 minutes to complete, and costs US$29.95 per capita.

Score reports are extremely detailed. They include a score on a scale of 20-80 for each of the five subsections of the test plus an overall score (Grammar and Vocabulary subscores, Reading Comprehension subscore, and Organisation and Voice & Tone subscores are weighted 60, 20 and 20 respectively in computing the overall score to reflect their relative importance for successful written communication). The score is accompanied by a description of overall capabilities associated with the particular score level. This is followed by an explanation of subskill scores, indicating what the particular area tested is covering (e.g., 'Organization reflects how well a candidate presents ideas and information in a logical sequence'), plus an individualized description of the capabilities of the candidate in that area at the particular score level achieved. These capabilities are then expanded with reference to the CEFR level descriptors to which the test was formally aligned. In addition, strategies and practices required for the test taker at this level to improve their writing skills are listed in the form of bullet points. The last section of the score report provides a measure of the candidate’s typing speed and accuracy to assist in score interpretation. It also indicates the approximate level on different scales (GSE and CEFR) to which the Versant score has been benchmarked.

Pearson offers guidelines and materials (in the form of samples of writing from the test at different score levels) to aid decision-makers in determining cut-scores between satisfactory and unsatisfactory competence for the intended purpose. A fee is charged for this service.

Equity
The instructions given to test takers for the 5 different test items are shown in a sample test paper, but there are no samples of actual test items available on the website. A demo can be requested by filling in the form on the website, and practice tests can be purchased at the cost of US$29.95.

A practice typing item at the start of each test allows candidates to familiarise themselves with the keyboard and measures typing speed and accuracy. If typing speed is below 12 words per minute and/or accuracy is below 90%, then it is likely that the candidate’s written English proficiency was
improperly measured due to poor typing skills. The test administrator should take this into account when interpreting test scores.
Appendix D. Samples of score reports of shortlisted tests

1. Academic English Screening Test/Post-entry Assessment of Academic Language

- Sample descriptors for AEST Writing (can be adapted for other contexts)

<table>
<thead>
<tr>
<th>AEST score</th>
<th>Criteria</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 (highest)</td>
<td>Form</td>
<td>You use a wide variety of sentence structures; your vocabulary is extensive and always used appropriately.</td>
</tr>
<tr>
<td></td>
<td>Fluency</td>
<td>You have shown that you are a competent writer. Your writing is fluent and well organized. The message can be followed effortlessly, the tone and style are appropriate to the task. Your essay shows a clear and logical progression of ideas.</td>
</tr>
<tr>
<td></td>
<td>Content</td>
<td>You provide sufficient ideas and evidence, although some examples might lack relevance.</td>
</tr>
</tbody>
</table>

Your writing shows that you are well equipped to cope with the writing demands of tertiary study.

2. Aptis
3. CaMLA English Placement Test (for institutions)
4. Diagnostic English Language Needs Assessment (DELNA) Screening

- Sample descriptors for DELNA Screening Writing

<table>
<thead>
<tr>
<th>DELNA band</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Your writing causes problems for the reader because links between ideas are inadequate or absent and the organisation is not clear. The data are not described clearly and some of your comments and ideas are confusing. You use an adequate range of sentence structures, but grammatical errors and limited vocabulary cause problems in your expression of ideas. Your writing shows that you are well equipped to cope with the writing demands of tertiary study.</td>
</tr>
<tr>
<td>6</td>
<td>Your writing needs attention. Inappropriate use of linking devices causes some strain for the reader. There are some problems with data description and the interpretation is not always adequate. You use a satisfactory range of sentence structures but are not always accurate in complex sentences. The errors that you make do not hinder communication significantly.</td>
</tr>
</tbody>
</table>

5. DIALANG (e.g., Listening)
6. Education First Standard English Test (Plus)

7. Oxford Online Placement Test (for Institutions)
8. Password

Password Test of English for University of XXXX

This is to certify that

Sample Student

achieved the following score in the Password Test of English for University of XXXX

taken on

March 15, 2011

Password Language Knowledge 5.5
Password Reading 5.0
Password Writing 5.0
XXXXX Listening 5.5
XXXXXX Speaking 4.5

Overall Score 5.0

9. Progress

Figure 5 Sample Progress Score Report for a single student
10. Versant English Placement Test
**Score Report**

**Overall Score**: 48

<table>
<thead>
<tr>
<th>Skill Area</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>40</td>
</tr>
<tr>
<td>Sentence Variety</td>
<td>60</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>55</td>
</tr>
<tr>
<td>Fluency</td>
<td>60</td>
</tr>
<tr>
<td>Pronunciation</td>
<td>60</td>
</tr>
</tbody>
</table>

**Description**

- Overall: The Overall Score of 48 represents the ability to convey a wide range of ideas in English speech. It is based on a weighted average of all sub-skills. Below are the scores for each sub-skill:
  - Sentence Variety: 60
  - Vocabulary: 55
  - Fluency: 60
  - Pronunciation: 60

**Candidate's Capabilities**

- Candidate converses effectively using a variety of words and structures, and uses common knowledge and frequently used expressions. Pronunciation is mostly clear, and it does not distract the listener. The level of English is consistent with a high intermediate learner.
12. Versant Writing

**Score Report**

**Versant Writing Test**

- **Overall Score**: 52

**Score Breakdown**

- **Grammar**: 50
- **Vocabulary**: 50
- **Organization**: 50
- **Voice & Tone**: 53
- **Readability & Clarity**: 57

**Description**

The candidate demonstrates awareness of language conventions and is able to maintain them consistently. The writing is clear and free of grammatical errors.

- **Organization**: The candidate organizes thoughts using a variety of strategies and maintains unity within the text. The writing is cohesive and logically structured.

- **Audience Awareness**: The candidate's writing addresses the intended audience effectively, using appropriate tone and style.

**Additional Notes**

- The raw score is 52, indicating proficient performance in the Versant Writing Test.

**Note**

Scores are based on a scale of 0 to 100, with higher scores indicating greater proficiency in English language skills. The Versant Writing Test evaluates a candidate's ability to write coherently and effectively for a variety of purposes.
SCORE REPORT

Test Identification Number: 12345678

Vocabulary
- Use a wide range of vocabulary that is appropriate to the context of the text and is accurate in terms of meaning and usage.
- Words used are relevant and add value to the message.
- Good command of vocabulary is evident, even in complex or technical contexts.
- Use of language is smooth and flows logically.

Strategic & Succinct
- Clear, concise, and direct expression that is effective for the intended purpose.
- Information is presented in a way that is easy to understand and remember.
- Use of language is appropriate to the audience and the context.
- Clear and logical structure of ideas.

TO IMPROVE, A CANDIDATE AT THIS TIME SHOULD CONSIDER:
- Practice reading faster or a wider range of texts to improve reading speed.
- Focus on higher-level comprehension by actively engaging with the text.
- Increase the variety of texts read to improve vocabulary and comprehension.
- Practice regular reading to maintain and improve reading skills.

SCORE REPORT

Text Identification Number: 12345678

Vocabulary & Accuracy
- The writer demonstrates a strong command of vocabulary and an excellent grasp of grammatical rules.
- Expression is clear, concise, and appropriate to the context.
- Use of language is appropriate to the audience and the purpose.
- Clear and logical structure of ideas.

Relational to Other Scores and Levels
- High scores in Vocabulary and Accuracy reflect a strong command of the language.
- The writer has a good understanding of the thematic concepts and is able to connect ideas effectively.
- The writer is able to manipulate language effectively to convey complex ideas.

NPSL
- Corresponding proficiency level: B2
- The Common European Framework of Reference for Languages (CEFR) is used to assess language proficiency.

CSE II
- Corresponding proficiency level in the German European Framework of Reference (CEF): B1
- The German European Framework of Reference (CEF) is used to assess language proficiency.

SAMPLE