Literacy Research and Instruction

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ulri20

Using a Five-Phase Applied Linguistics Design to Develop a Contextualized Academic Literacy Placement Test for Pre-University Pathway Students

Jonathan H. Green, Charmaine Davis, Marcus Harmes, Kate Judith & Albert Weideman

To cite this article: Jonathan H. Green, Charmaine Davis, Marcus Harmes, Kate Judith & Albert Weideman (2025) Using a Five-Phase Applied Linguistics Design to Develop a Contextualized Academic Literacy Placement Test for Pre-University Pathway Students, Literacy Research and Instruction, 64:2, 229-255, DOI: 10.1080/19388071.2024.2340031

To link to this article: https://doi.org/10.1080/19388071.2024.2340031

Published online: 25 Apr 2024.

LITERACY RESEARCH AND INSTRUCTION
2025, VOL. 64, NO. 2, 229–255
https://doi.org/10.1080/19388071.2024.2340031

Routledge
Taylor & Francis Group

Using a Five-Phase Applied Linguistics Design to Develop a Contextualized Academic Literacy Placement Test for Pre-University Pathway Students

Jonathan H. Green [a,b], Charmaine Davis [a], Marcus Harmes [a], Kate Judith [a], and Albert Weideman [b]

[a]UniSQ College, University of Southern Queensland, Toowoomba, Australia; [b]Faculty of the Humanities, University of the Free State, Bloemfontein, South Africa

ABSTRACT

For students who have faced previous educational disadvantage, academic literacies are key to access and participation in higher education. Reliable placement in academic literacies courses engages these learners in appropriately pitched learning activities, ensuring optimal learning while developing their confidence and motivation. In this context, we aimed to develop a reliable assessment of students' academic literacy knowledge to ensure accurate and context-sensitive placement in one of two foundational academic literacy courses. These courses were situated in a pathway program in a regional Australian university that aimed to broaden participation in higher education for individuals from under-represented groups and educationally disadvantaged backgrounds. In this article, we describe, sequentially, the use of a five-phase applied linguistics design to develop the Academic Literacy Level Test (ALLTest). The phases represent the underlying principle of allowing pragmatic considerations to precede (but not preclude) theoretical ones, and consist, in turn, of (1) identifying the language challenge, (2) applying technical imagination and knowledge, (3) devising an initial (and iteratively subsequent) solution, (4) providing a theoretical justification, and (5) revising or finalizing a blueprint. We demonstrate how this process resulted in an organizationally efficient, context-specific, valid and reliable test of academic literacy.

KEYWORDS

Academic literacy; placement testing; widening participation in higher education; test development

Introduction

Arguments for producing a good fit for learning for students are compelling and manifold. Among these reasons is that effective learning occurs when the content and instruction are appropriate to the developmental level of the students concerned; in other words, it is neither too easy nor too difficult. In this respect, for example, Vygotsky's (1978) notion of the Zone of Proximal Development (ZPD) has been enduring: the idea is that learning is best facilitated when it is set at a level at which a learner may not be able to complete the learning task independently but may be able to do so with some guidance from a capable instructor or peer. In language development, the notion of an ideal zone for learning has been adapted by Krashen (1987) in his input hypothesis, which emphasizes the importance

CONTACT Jonathan H. Green Jonathan.Green@unisq.edu.au UniSQ College, University of Southern Queensland, 487-535 West Street, Darling Heights, Toowoomba, QLD 4350, Australia

of input in the target language that may be *just* short of being fully comprehensible to the learner. In this respect, academic literacy development in individual learners is akin to additional language development: for some students, the discourses and practices of academia are not unlike negotiating a foreign culture and language (e.g., Finnegan & Merrill, 2017).

Concepts such as the ZPD and input hypothesis are illustrative of the notion of the existence of a "Goldilocks zone" for learning – a place where learning is neither too difficult nor too easy, but "just right." Placement testing aims to place groups of learners as suitably as possible in this zone (Noble et al., 2003), particularly where an institution offers multiple levels of instruction – a progression – within a domain, such as second-language learning, numeracy, or academic literacies.

Ideally, academic literacy placement testing is not merely diagnostic of individual cognitive functioning (Beleche et al., 2012); it recognizes literacies as "a range of social and cultural practices around reading and writing in particular contexts" (Lea, 2017, p. 147) and, in the case of academic literacies, best situates the learner for positive and meaningful interaction within the academic learning environment.

Appropriate placement has other benefits important in a widening participation context. It:

- ensures that students are treated equitably, according to their respective needs and developmental levels (Kirst, 1998).
- assists student success by giving them appropriate advice on preparation for their chosen university pathways (Morante, 1989); and
- provides an opportunity, from a non-deficit perspective, of evaluating courses to ensure an appropriate fit with students (Beleche et al., 2012).

For students who have faced barriers to academic attainment, addressing academic literacy is key to providing them with access to university study (Baker & Irwin, 2015a) and equipping them for success. Reliable context-appropriate placement is, therefore, crucial.

This paper presents a case study of a contextualized placement test development for a pre-university pathway program that included people from under-represented groups and marginalized academic backgrounds. It describes how a team at the Tertiary Preparation Program (TPP) at the University of Southern Queensland, Australia, used Weideman's (2019a) five phases of language intervention design to improve existing placement practices by creating a valid and reliable placement test of academic literacy to help inform students' enrollment in one of a progression of two academic literacy courses.

The paper describes the widening participation and institutional context of the test and then outlines sequentially the development of the resulting functionally based three-tiered Academic Literacy Level Test (ALLTest) in terms of each of the five phases of applied linguistics design described by Weideman (2019a).

We demonstrate how considerations in each of these phases resulted in a reliable, comprehensive, contextualized placement test that was engaging and accessible while possessing the necessary discriminatory power to effectively determine the appropriate course for students starting their pre-university journey.

The context for development

Academic literacy placement testing in the higher education context

In Australia, a program that seeks to broaden participation in higher education is often referred to as an "Enabling course." It is defined as "a course of instruction provided to a person for the purpose of enabling the person to undertake a course leading to a higher education award" (Australian Government, 2017, p. 384). Such programs are also variously referred to as bridging, preparatory, foundation or pathway programs (Hodges et al., 2013) and are provided to students fee-free (Australian Government, 2017).

Testing upon entry is employed for a variety of purposes in Australian Enabling programs. Baker & Irwin (2012) determined that approximately half of Enabling programs required students to complete academic literacy testing upon application, while Davis et al. (2023) determined that two out of a panel of nine Enabling programs used academic literacy testing as a determinant for entry, with a further three using testing for course placement and support purposes. However, there has been little research in Australia on the development, implementation and efficacy of pre-entry academic literacy placement testing in the pathway programs; instead, there has considerable focus on assessing students' English language proficiency upon application for university undergraduate placement. However, Sebolai (2016) argues tests that focus on assessing English language proficiency are not effective in assessing the preparedness of students to meet the academic discourse demands of university study.

As most universities in Australia do not directly assess the academic literacy proficiency of prospective students, and instead rely on school achievement or other measures to assess admission applications, there has been little impetus to research and develop effective academic literary testing processes. However, reliance on traditional school achievement measures for university entry does not adequately account for students who have experienced educational disadvantage during their schooling experience, and this has prompted research internationally in the use of alternative measures for determining academic literacy preparedness for undergraduate study (Cliff & Hanslo, 2010). In the U.S., standardized testing, including SAT and ACT for college entry, has been widespread, but has also been a contested phenomenon. In recent years, many colleges have opted to remove "gatekeeping" test scores as a compulsory component of college application, citing the impact of the COVID-19 pandemic, declining admissions and concerns over test bias (Allensworth & Clark, 2020), instead relying on high school performance and other measures to assessment college applications.

The efficacy of academic literacy tests to determine course placement on an institutional level, particularly for students subsequently enrolled in developmental education programs, has also been subject to scrutiny in the US context. Of particular focus is retention and success for students in these programs, the challenges of developing appropriately contextualized tests that align with the curriculum and are adequate predictors of student performance outcomes (for example, Cooper et al., 2019; Valentine et al., 2017).

In the South African context, moves toward massification of higher education have similarly prompted the development of academic literacy placement mechanisms beyond reliance on secondary school outcomes. Weideman (2019a) argues that close alignment between design constructs of academic literacy placement testing and consequent instruction is critical for successful academic literacy interventions in higher education. These

principles are reflected in the Test of Academic Literacy Levels (TALL), developed by Weideman (2021) and now used amongst a consortium of South African universities to inform admission and course placements.

Lessons from the literature

It is difficult to identify a unified approach in Australia to the type of post-enrollment placement testing contemplated in the development of the ALLTest. Nevertheless, the literature that has emerged from placement testing from US and Canadian developmental education (which is cognate with Enabling education in Australia) and from English language proficiency testing is instructive.

Criteria for placement testing. Throughout this development, the imperatives of fairness, validity and reliability have emerged as key principles. These tenets are expressed through the literature in several criteria, including:

- Having a clear purpose and goals (e.g., Hodara et al., 2012; Venezia et al., 2010);
- Achieving fairness by avoiding cultural bias and ensuring the test is available and accessible to a diverse group of students (e.g., Barnett et al., 2018; Fulcher, 1997; Golder, 2006);
- Developing test items that reflect the knowledge, content and skills essential for success in university courses (Barnett et al., 2017; Noble et al., 2003);
- Aligning the test format and structure with the knowledge and skills being assessed and, in so doing, considering a balance in different question types (e.g., short answer, multiple choice, etc.) and the overall length of the test (e.g., Burdman, 2012; Dimova et al., 2020);
- Developing a reliable scoring system and clearly defining the interpretation of test scores. Integral to this is considering providing diagnostic information to students, teaching staff, and university administrators to guide targeted interventions (Elliot et al., 2012; Murphy & Yancey, 2009)

Further afield, these criteria are echoed in the literature on English language and academic literacy placement (for an international perspective, see, for example, Elder & Read, 2015; Weideman et al., 2020)

Limitations of placement testing. In addition to the key criteria outlined above, the literature on placement testing in developmental contexts reveals some key limitations, most of these related to placement testing being a single, one-time measure. The tests are, according to critics:

- Weak predictors of student success in undergraduate courses (e.g., Barnett et al., 2018; Burdman, 2012);
- Unable to provide insight into a student's potential for development and improvement over time (e.g., Scott-Clayton et al., 2014);
- Not reflective of the changing landscape of higher education, especially in terms of evolving digital multimodality and advancements in generative artificial intelligence (e.g., Mokher et al., 2023); and

- Unable to provide a comprehensive perspective of the preparedness of students' overall readiness for university study (e.g., Armstrong, 2000) particularly where typical English and mathematics assessments are concerned (Burdman, 2012).

According to Burdman (2012), many of these limiting factors are problematic because placement tests are often high stakes (they are often used as gatekeeping measures): premising students' entry into undergraduate courses and programs based on a "myopic" measure that is a weak predictor of success is tenuous. It may also be inequitable for students who enter university through non-traditional pathways.

Burdman further recognizes that many of these challenges may be addressed by, for example, "downplaying the tests," seeking multiple measures, establishing informed self-placement, and devising assessments that are aligned with curriculum and required knowledge and skills.

The ALLTest in the Australian widening participation context

Of critical consideration in determining the purpose, design, development and implementation of the ALLTest is not only the institutional context but the broader international context of widening participation. The University of Southern Queensland (UniSQ) is an Australian regional university that has one of the most diverse student bodies of any Australian university, with most of its approximately 25,000 students (University of Southern Queensland, 2022, p. 70) regionally based. Many students are from identified equity groups, including from low-socioeconomic, non-English speaking, first-in-family to attend university, and Indigenous backgrounds, along with a significant number of incarcerated students undertaking study from prison facilities across the nation. This institutional diversity reflects a response to international efforts to expand and widen participation in higher education (Burke, 2013).

The widening participation agenda is in part manifested through UniSQ College, the remit of which includes the delivery of pathway programs that provide preparatory study and access to undergraduate programs. Currently, almost 12% of students admitted to undergraduate programs at the university have articulated from UniSQ College's widening participation programs.

In Australia, Enabling courses (see previous section) have become increasingly popular as universities have sought to attract a more diverse range of students. Approximately 50% of students undertaking these programs nationally are from identified equity groups (Lomax-Smith et al., 2011), compared with 30% of all domestic undergraduate enrollments. Pathway programs thus play a critical role in the widening participation agenda (McKay et al., 2018). The increasing size and impact of pathway programs at UniSQ and nationally reflects a global trend in similarly-tasked programs internationally (Agosti & Bernat, 2018; Neghina, 2015).

The Tertiary Preparation Program (TPP) is a keystone Enabling program offered by UniSQ College. The program, among the largest of its type in Australia (Department of Education Skills and Employment, 2020), has experienced a 100% increase in student numbers since 2014. As annual enrollments had moved beyond 3,000 students at the time of the development of this test, there was a clear imperative to develop not only

more nuanced course placement outcomes, but more efficient administrative processes to continue to provide a high level of service to students.

Typically, Enabling programs report an attrition rate of approximately 50% (Hodges et al., 2013; Klinger & Murray, 2012), and improved retention and outcomes for students have been a sustained focus of research in the field. While a decision not to complete a university program can be based on the complex interaction of social, academic, and individual personal factors (Christie et al., 2004), much research into reasons behind attrition in Enabling programs concludes that personal circumstances (Lisciandro & Gibbs, 2016) and a sense of connectedness to and support by the university community are significant factors (Bennett et al., 2018; Chesters & Watson, 2016; Willans & Seary, 2018). Well-designed placement testing may help students to connect by easing them into the sociocultural context at university; it also provides them with targeted support through informed enrollment into courses that appropriately scaffold and develop their students' existing academic literacy knowledge – knowledge that is critical in all university studies (Grebennikov & Shah, 2012), and even more so for Enabling education students.

While there is no formal Australian framework or formalized standards governing them, common learning outcomes (National Association of Enabling Educators of Australia, 2019) underpin many of these programs. Typically, program learning objectives include the development of numeracy, academic and digital literacies and study management skills, alongside introductory knowledge of academic disciplines relevant to individual student aspirations (Baker & Irwin, 2015b; Brett et al., 2016; Hodges et al., 2013). A focus is addressing the academic literacies (Baker & Irwin, 2015a) necessary for successful tertiary study. These include the ability to find, retrieve, understand, analyze, synthesize, and critically evaluate information in different formats, along with the communication and critical thinking skills required to engage with and produce academic texts. There is also a recognition that academic literacies are not abstract, unitary skill sets, but that they develop and are used by participants interacting in a certain sociocultural milieu (e.g., Henderson & Cunningham, 1994) – in this case, the academic context of the university.

UniSQ's TPP exemplifies these broad characteristics and aims. Furthermore, the program is open to all students and does not have minimum entry requirements, resulting in a very diverse range of knowledge and experience amongst the student cohort. Consequently, the program has different entry levels, including two academic literacy courses. Students applying for the program can undertake academic literacy and numeracy placement tests to ensure they are appropriately placed. Previously, these placement tests were paper based, accessed by prospective students through the university's website, completed and returned to the College for marking by academic staff. An enrollment recommendation was then made to students, who were subsequently enrolled in courses by college administrative staff.

The development and implementation of the ALLTest set out to address significant limitations inherent in this process, particularly in the case of the academic literacy test, then termed the *Communication test*, or *C-test*. Firstly, increased student numbers meant managing the C-test process became increasingly unsustainable. Moreover, the validity and reliability of the existing placement test were questionable. The C-test was originally based on an English language assessment, rather than having a discrete focus on academic literacies, and was, therefore, not as effective as a predictor of academic literacy knowledge and ability as it could have been. The untimed and unregulated implementation of the test also raised questions about the authenticity of answers and meant appropriate course

placement was at times compromised. Furthermore, test papers were marked holistically, which failed to provide adequate data about the specific existing skills and knowledge of new students. There was a need for an operationally streamlined process to implement a test that directly targeted selected academic literacy constructs and could be delivered in a more controlled manner. The development of a new test also provided the opportunity to gather detailed data that could inform course development to ensure alignment with the learning needs of the student cohort.

The five phases of applied linguistics design

In designing language interventions in educational institutions, Weideman (2019a) asserts that alignment between assessment and language instruction is critical, and that such interventions, to be effective, require support by institutional policies. To achieve such alignment, Weideman describes a five-phase process of applied linguistics design for tests of academic literacy. He argues that this five-phase model best accommodates "the various possible combinations of policy, curriculum (instructional), and assessment interventions designs that one may find in actual institutional processes" (p. 34).

Alignment is also a consideration in determining test validity; a model that could achieve alignment would therefore also address such validity. In discussing the validation and validity of language tests, Weideman (2019c; also Weideman & Deygers, 2024) argues that the conventional practice of associating the concept of validity purely with interpretations of test scores is problematic. Instead, because language testing occurs within the domain of applied linguistics, the design of a language test – and, by extension, an academic literacy test – should be subject to the same criteria as those that apply to all applied linguistic artifacts. Such design recognizes the primacy of "the technical dimension of our experience" (p.4) and that validity should be multi-faceted, including considerations such as the consistency of measurement, a clear notion of what is being tested – the theoretical rationale of the test – and, thus, the interpretability of test results.

In addition to alignment and validity, it was important that we considered the technical possibilities of our context (e.g., the affordances and limitations of the Moodle-based learning management system employed by the university) and our analytical purposes (e.g., appropriate placement). The five-phase development model (Figure 1) was suitable as guide in this regard, too, as it recognized and accommodated in its initial phases these pragmatic elements and a consideration of the contextual possibilities.

Moreover, the model recognized that design is more often iterative than linear. It thus accorded with our primary purpose, which was not to develop theory but to find a solution to a practical pedagogical and operational challenge. This is not to argue that a theoretical underpinning was not essential: only that the initial solution would be impelled by pragmatic considerations, and that this solution would then be considered in terms of applicable theory to evaluate its conceptual justifiability. Adjustments would be made, if required, to derive a final blueprint.

Corresponding with our overall purpose and the considerations outlined above, the five-phase design (see Figure 1) is as follows:

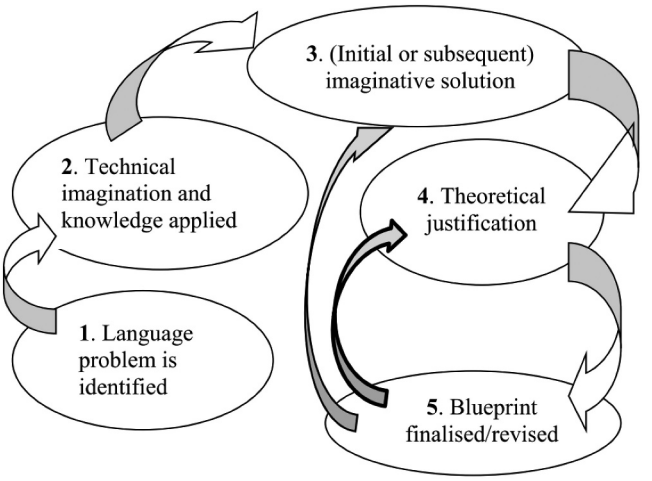

Figure 1. The five phases of applied linguistics design (Weideman, 2019c).

(1) The "language problem" is identified and articulated;
(2) The collective technical imagination and theoretical knowledge of the designers, as it potentially applies to the "language problem," is brought into consideration;
(3) An initial imaginative solution to the "problem" is formulated;
(4) A theoretical justification is sought and designed for the proposed solution; and
(5) The solution is reviewed in terms of its adequacy and appropriateness and redesigned if necessary.

Test development

In the section that follows, we outline the development of the resulting placement test, the Academic Literacy Level Test (ALLTest), through each of these phases.

Phase one: identifying the language "problem"—placing pre-university pathway students appropriately and efficiently in a language literacy course

The "problem" that was to be addressed here was the pedagogical and organizational imperative to place students appropriately and reliably in one of two academic literacy courses with the Tertiary Preparation Program at UniSQ – we emphasize that "problem" in this sense is used in consistency with the labeling of the first design phase and is, by no means, intended to convey a deficit view of students, their knowledge or ability. It is important to briefly but early acknowledge the implications of language as participating in a deficit discourse. When used in this discussion, however, we use it with reflexive awareness of its implications and connotations. In this sense, "problem" represents the institutional challenge to provide the best possible student learning experience within the learning context, recognizing the centrality of academic literacy in providing students with access to university study and equipping them for the subsequent pathway to graduation.

Phase 2: applying technical imagination and knowledge to the academic literacy placement challenge

The design of the ALLTest for students entering the University of Southern Queensland proceeded from the premise that for such an assessment to be valid (Weideman, 2019b, 2019c), its construct first had to be clearly articulated. It took its cue from the following definition of academic language:

> Academic discourse … includes all lingual activities associated with academia, the output of research being perhaps the most important. The typicality of academic discourse is derived from the (unique) distinction-making activity which is associated with the analytical or logical mode of experience. (Patterson & Weideman, 2013)

It follows that one, therefore, needs to articulate that construct further in such a way that justice is done to the typicality of academic discourse. In language testing, this typicality is often expressed as a congruence between the kinds of language tasks associated with the domain (academic language use) and the kinds of tasks in the assessment instrument.

Attending to such congruence validates the assessment by demonstrably being in alignment with the domain being tested and fulfills to a considerable extent the requirement for a test design to be technically useful. As will be noted below, all but one of the test tasks selected for use in the various subtests have this kind of authenticity; in the case of the one exception, the last subtest on *Grammar and text relations*, an argument can be put forward to conclude that it tests certain sub-skills that are often used in working with academic language, such as looking for the right term, finding the correct preposition, or selecting the appropriate grammatical feature or construction.

Having identified the construct ("academic language"), the main challenge then is to operationalize it. Doing that inevitably involves a measure of reduction. Such reduction is subject, nonetheless, to the development of a theoretically justifiable set of components of academic language. What is to be assessed is broken up into several components of the ability to use academic language, which ability can be referred to as academic literacy. Those components are required to derive from perspectives on language that are theoretically defensible (as discussed below). In the current case, those perspectives are the functional views of language associated with the work of Hymes (1972), Habermas (1970) and Halliday (1978). Without such a clear and theoretically defensible articulation of what the ability is that is being assessed, the technical meaningfulness of the test would be in jeopardy, for the interpretation of the results of the measurement must be done with reference to what has been measured. In developing ALLTest, the formulation of these components of academic literacy, as shown in Table 1, was drawn from Weideman's (2021) functional approach to academic literacy.

Of course, not all these components would carry equal weight or would be entirely relevant in tests for students entering university-level study for the first time; a selection had to be made from them that was appropriate for the ability that new entrants are supposed to have to handle academic language. To test the language ability that has been articulated from such selection, the components can then be further matched up with task types that might be used to assess the level of mastery of academic language. In the case of ALLTest, such matching up was accompanied by the further step of assigning different weightings to the various tasks in the three levels or tiers of the administration of the test, the justification for which we return to below,

Table 1. Formulation of components of academic literacy (Weideman, 2021).

The ability to use language for academic purposes is a functional ability to:	
understand and use a range of academic vocabulary as well as content or discipline-specific vocabulary in context	distinguish between essential and non-essential information, fact and opinion, propositions and arguments, cause and effect; and classify, categorise and handle data that make comparisons
interpret the use of metaphor, idiom and non-literal expression in academic language, and perceive connotation, word play and ambiguity	see sequence and order, and do simple numerical estimations and computations that express analytical information, that allow comparisons to be made, and can be applied for the purposes of an argument
understand and use specialised or complex grammatical structures correctly, also to handle texts with high lexical diversity, containing formal prestigious expressions, and abstract/technical concepts	systematically analyse the use of theoretical paradigms, methods and arguments critically, both in respect of one's own research and that of others
understand relations between different parts of a text, be aware of the logical development and organisation of an academic text, via introductions to conclusions, and know how to understand and eventually use language that serves to make the different parts of a text hang together	interact with texts both in spoken discussion and by noting down relevant information during reading: discuss, question, agree/disagree, evaluate and investigate problems, analyse
understand the communicative function of various ways of expression in academic language (such as defining, providing examples, inferring, extrapolating, arguing)	make meaning of an academic text beyond the level of the sentence; link texts, synthesize and integrate information from a multiplicity of sources with one's own knowledge in order to build new assertions, draw logical conclusions from texts, with a view finally to producing new texts, with an understanding of academic integrity and the risks of plagiarism
interpret different kinds of text type (genre) with a sensitivity for the meaning they convey, as well as the audience they are aimed at	know what counts as evidence for an argument, extrapolate from information by making inferences, and apply the information or its implications to other cases than the one at hand
interpret, use and produce information presented in graphic or visual format in order to think creatively: devise imaginative and original solutions, methods or ideas through brainstorming, mind-mapping, visualisation, and association	interpret and adapt one's reading/writing for an analytical or argumentative purpose and in light of one's own experience and insight, in order to produce new academic texts that are authoritative yet appropriate for their intended audience

Note: components are not taxonomically arranged.

after considering several further key contextual and functional considerations that underpinned the test design.

Firstly, logistical and administrative ease was important. The test would need to be functional within the institutional context and limit the administrative burden. With large student numbers needing to obtain their test results quickly so that they could proceed with correct self-enrollment, any unnecessary administrative actions had to be avoided. As a result, the test would need to aim toward effortless integration within the online learning environment of UniSQ College and maximize the capacity for automation. These considerations pointed to the use of a multiple-choice format.

Secondly, several design considerations were related to minimizing the cognitive instructional burden for the students. Instructional simplicity was needed to avoid confusion or stress that may lead to failure to complete, and the design would need to include a mechanism to enable the test to be reliably employed to minimize redundancy and optimize efficiency.

Thirdly, accountability was paramount. The role of the test would be to provide accurate advice to students to guide them to make the most appropriate decisions for their self-enrollment, rather than gatekeeping entry. Nonetheless, we needed the test to be rigorous enough to support us in providing the best advice to students.

Fourthly, the test needed to be fit for purpose. Very few international students enroll in the program, so the communication courses in the TPP program aim to prepare domestic students for successful academic communication at university, specifically in undergraduate programs.

The components of the relevant academic communication proficiencies outlined above are not specific English language skills but rather include such critical thinking competencies as *being able to comprehend and interpret written information*, *evaluate the quality of argument*, *identify logical flaws*, *interpret graphic information*, *sequence an argument*, and *use evidence to support an argument*. The test was therefore neutral regarding any focus on discrete language skills (Weideman, 2021); that is, it would need to avoid selecting students based on specific grammar, spelling, or vocabulary skills, or any attempt to divide language competence between the skills of listening, speaking, reading, and writing. Rather, it should aim to select based on the relevant communicative functions that were outlined above as components of the construct.

Phase 3: the initial imaginative solution

To meet the fourfold objectives of operational ease, minimized cognitive load on students, accountability, and fitness-for-purpose, we workshopped some initial design decisions.

Firstly, we chose a three-tier structure with subsequent tiers increasing in their level of difficulty. We anticipated that this structure would increase administrative efficiency and reduce the time and stress load for students by selecting out bands of students whose results at each tier indicate further tiers of testing are not needed, and by positioning the more intensive evaluative focus on the borderline student group. Furthermore, we developed the first and second tiers to be computer assessable to minimize marker time. We introduced a longer written response of 450 words and three paragraphs for the third tier. We expected fewer responses here, so manually marked assessment could be accommodated in this tier.

Secondly, the team chose assessment tasks and developed them based on how they matched the components of the construct (see Table 1). For example, the *Vocabulary* subtest would measure lexical knowledge (in this case, as defined by Coxhead, 2000), while a *Scrambled text* subtest would assess the sub-ability to see relations between various parts of a text, to organize text, to recognize and use a communicative function, and so on. Some subtests, like *Text comprehension*, would enable one to assess not only the ability to define, compare, contrast, make distinctions, infer, extrapolate, and conclude, but also on occasion test vocabulary, text relations, comparing text with text, and so on. See Table 2 for examples of items that we developed for the subtests.

We selected tasks for the subtests that make up the three tiers of the test and weighted them for each tier as shown in Tables 3 and 4. The subtests selected for each tier were selected from experience with similar tests in other environments. In these, for example, the *Scrambled text*, *Vocabulary* and *Grammar and text relations* subtests were the ones whose scores correlated most closely with the overall test scores of a longer test and were deemed appropriate, therefore, for use in Tier 1. We designed this tier to
be employed either to identify the students whose ability was already at the required level and thus did not need to have to complete a second, longer test of language ability, or to

Table 2. Sample items for selected sub-tests of the ALLTest (prior to digital conversion).

Scrambled text

The sequence of the sentences in the following has been altered. Say what the correct order is by answering questions 1–4.

Threatened woodlands - Queensland's Brigalow Belt

A. This unusual, long-lived acacia with its dark, fissured bark and distinctive silver leaves forms dense woodlands, home to unique and threatened plants and animals.

B. Before clearing, brigalow-dominated ecological communities covered an estimated 7.5 million ha within the Brigalow bioregion, but those vast brigalow woodlands are no longer here.

C. Although the region contains diverse ecosystems, from dry vine scrub to grasslands, it is named after the species of tree that once dominated: the brigalow (*Acacia harpophylla*).

D. Extending over an area of 36.4 million hectares from Townsville down into New South Wales, Queensland's Brigalow Belt is among Australia's most significant biodiversity hotspots.

[adapted from *The Conversation*, 6 May 2016]

1. Which sentence did you put **first**?	A	B	C	D
2. Which sentence did you put **second**?	A	B	C	D
3. Which sentence did you put **third**?	A	B	C	D
4. Which sentence did you put **fourth**?	A	B	C	D

One-word vocabulary

Choose the best possible answer from the list of options:

Mountain Ash forests, dominated by giant *Eucalyptus regnans* trees, _____ moist climates and deep soil to maintain their health.

A. refer

B. require

C. request

D. refuse

Two-word vocabulary

Choose the best possible answer from the list of options:

Vast tracts of rainforest have already been logged and the wildlife has _____ been hard pressed to _____ in these circumstances.

A. conclusively … dedicate

B. confusingly … order

C. contagiously … manage

D. consequently … survive

ister and test type

The sentences below are examples of different text types, such as advertisements, instruction manuals, academic textbooks and the like. You must match an item from the first set (31–35) with an item from the second set (A-E). For example, if you think that the language of 32 comes from the same text type as A, then mark 32 A as your answer.

31. Yet so far only one modern scholarly study, T.T.F. Plucknett's *Legislation of Edward I* (1962), has attempted to look at the period.

32. Motorists who push to beat the amber light are now being stung with a $400 fine and three demerit points, the same penalty for running a red light.

(*Continued*)

Table 2. (Continued).

Scrambled text

33. It was a derelict place full of webs and bird nests and flickering shadows and the eeriness of it distracted us from our awkwardness.

34. Licensed 1892 The Spotted Cow Toowoomba's Famous Steakhouse & Bar 46,393,264

35. Serious interactions have been reported after the use of high dose methotrexate with PANAMOR AT-50.

A. The pounding of my own heart reverberated in my ears with such force I couldn't hear the cicadas screaming up above.

B. Sunset Superbowl – Toowoomba – Open from 9 AM every day – 531 South Street – www.sunsetsuperbowl.com.au

C. Acting Senior Sergeant Loveday said police needed to adopt a common sense approach when dealing with potential offenders.

D. Increased creatine kinase (CK) levels (from skeletal muscles) as well as muscle tenderness have been frequently reported.

E. To show that there were important developments in his reign, this paper will conclude by looking at one such legal development.

Grammar and text relations

In the text below some words have been deleted. First read through the whole text, then answer the questions that follow.

Ecosystems

(source http://www.softschools.com/language_arts/reading_comprehension/science/66/ecosystems/)

WHAT IS NEEDED TO SURVIVE? How do livings things interact? Is there enough food, clean water, comfortable temperature, shelter? All of the answers to questions are found in the ecosystem. An ecosystem is all living things non-living things and how they interact each other. Ecology is the study of how these things interact with each other order to survive.

An ecosystem may be very small like a backyard or a pond. Or can be large like a prairie, a desert, or a rainforest. There freshwater ecosystems … [*full article continues; excerpts and items below are drawn from the full article.]*

In the following, you have to indicate both the possible *place* where a word may have been deleted, and which *word* belongs there.

Regardless of where they are located or $_{\text{Example 5\&6}}$ i how small or ii large they iii be iv , all $_{\text{Example 7\&8}}$ i ecosystems ii are iii made up iv the same parts.

Example 5: Where has the word been deleted?	**Example 6: Which word has been left out here?**
A. At position (i).	A. of
B. At position (ii).	B. really
C. At position (iii).	C. together
D. At position (iv).	**D. may**
Example 7: Where has the word been deleted?	**Example 8: Which word has been left out here?**
A. At position (i).	A. these
B. At position (ii).	**B. of**
C. At position (iii).	C. often
D. At position (iv).	D. with

Notes. 1. Samples are for the purpose of illustration only and are drawn from non-included items of the ALLTest; no claims are made as to their discriminatory power, level of difficulty or validity. 2. For *Text comprehension* and *Interpreting graphs and visual information* examples, see Figure 2 *Moodle implementation of the ALLTest*..

Table 3. Tier 1 subtests and mark allocation.

Subtest	Function(s) tested	Marks
Scrambled text 1	Complex grammar; text relations; inference; extrapolation; synthesis of information; constructing an argument	4
Scrambled text 2		4
Vocabulary (one-word)	Vocabulary and metaphor	6
Vocabulary (two-word)		4
Grammar & text relations	Complex grammar; text relations; communicative function	12
Total		30
Duration		30 minutes
Results in two categories:		
Below 80% - Take Tier 2 test		
Above 80% - No need to take any further test		

Table 4. Tier 2 subtests, mark allocation and bands.

Subtest	Function(s) tested	Marks
Scrambled text	Complex grammar; text relations; inference; extrapolation; synthesis of information; constructing an argument	5
Vocabulary (one-word)	Vocabulary and metaphor	8
Vocabulary (two-word)		6
Interpreting graphic & visual information	Text type, including visually presented information; essential/non-essential information; sequence and numerical distinctions; identifying relevant information and evidence	10
Register & text type	Text-type/register awareness	5
Text comprehension	Vocabulary and metaphor; communicative function; text type; essential/non-essential information; identifying relevant information and evidence	30
Grammar & text relations (2+2+12)	Complex grammar; text relations; communicative function	16
Total		80
Duration		75 minutes
Results in risk bands:		
Band 1: Very high risk		
Band 2: High risk		
Band 3: Borderline - Write Tier 3 test		
Band 4: Less risk		
Band 5: Little to no risk		

provide all students who wished to prepare for the Tier 2 test with a chance to familiarize themselves with the format. For those students scoring below 80% in Tier 1 (see Table 3), taking Tier 2 was deemed to be essential; Tier 2 is a longer test and, therefore, able to give a more reliable reading of the language ability of those tested.

In each case, we carefully selected the texts used in these tiers based on their reading ease and grade level, and these measures of difficulty increased gradually over the tiers. For the Tier 3 test (see Table 5), all the texts used in the first two tiers were available, as well as an additional text that dealt with the same theme and allowed the development of a written argument.

Thirdly, a scoring rationale capable of justifiably identifying the cut-off scores was needed. We estimated these scores and flagged them for further refinement once the tests had been piloted. The intention was that students scoring above 80% in Tier 1 would not be required

Table 5. Tier 3 mark allocation.

Subtest	*Marks*
Making academic arguments (450 words, three paragraphs) [Add additional text to prior texts]	
Total	20
Duration	60 minutes
For second/third chance attempts	
For those with Category 3 results and incapacitated test takers of Tier 1 & 2 tests	

to proceed to the next tier and could be advised to enroll in the higher of the two academic communications courses.

For Tier 2, we identified five risk bands based on the students' scores (see Table 4). "Risk" in this case corresponds to Gordon's (2006) argument that placement assessment should be designed to provide an estimate of a student's ability to attain the basic demands of the course; thus "risk," in this context, refers to the potential that the student would not meet such demands and thus fail the course. Those in Bands 1 and 2 (*very high risk* and *high risk*) would be advised, therefore, to enroll in the appropriate foundational-level course. Those with Band 3 results (*borderline*) would be eligible for a second chance, in the form of the Tier 3 test, as would those who were unable to complete earlier tiers. Students in Band 4 and 5 were to be regarded as having *little* or *no risk* and could be provided with advice to enroll in the more advanced course.

Ensuring a calibration methodology for test results so they would enable sound advice for students was fundamental to the design process from the start. A well-established institutional placement test, the C-test was already in use. This older test was based on a 450–500–word written response to a prompt, requiring evidence from a provided selection of texts and some short – answer questions. While, as indicated in a previous section, the C-test was operationally inefficient and of questionable validity and reliability (being based, for example, on standardized English language proficiency rather than a sound construct of academic literacy), it was available as a benchmark, so for the first two intakes, students undertook both tests. This allowed for the calibration of the new ALLTest. In addition, the older C-test would be available to be modified for use as Tier 3 in the new test.

Phase 4: Theoretical Justification

While the initial design incorporated both contextual knowledge and initial theoretical considerations, we revisited these to evaluate their applicability to the initial design and to further articulate the theoretical justifications for the design decisions that we made.

Because we recognized the importance of academic literacies in providing access to and equipping students for success at university, it was important to adopt a view of academic literacies that was grounded in the reality of what students realistically needed to know and commonly use in negotiating an academic setting. It followed that the test would, therefore, reflect this view of academic literacies through its constructs and underpinning theoretical justification. To provide this justification we evoked, at least provisionally, the seminal ideas of Habermas, Hymes and Halliday in our initial design ideas.

Together, these theorists provide a sound justification for a communicative, sociocultural underpinning to language development. Hymes (1972), in the first instance, provides a rationale for the design approach that we have taken; he argues, for example, that "work motivated by practical needs may help build the theory we need" (p.1). Correspondingly, the five-phase design is motivated first by a practical need for an efficient, reliable, and fit-for-purpose placement test, the development of which then evokes reflection on a theoretical justification. This is unlike conventional deductive approaches to research and underlies, too, the somewhat unconventional but intentional approach taken in this discussion to introduce the pragmatic considerations of the test development prior to outlining the theoretical justification.

Hymes dismisses Chomsky's (1965) focus on an abstract system of language, which posits an "ideal language user." This ideal language user was a theoretical construct in Chomsky's theory of generative grammar that represented a speaker's or writer's complete and perfect knowledge of their language, including its grammatical rules, structures, and infinite generative capacity, free of the variables and limitations of actual language performance. Instead, Hymes advocates a communicative perspective that focuses on a language user's knowledge of the patterns of use within a specific social context (Cazden, 2011). In sum, Hymes's formulation of communicative competence as the knowledge needed to use language in a particular social context is compatible with the holistic approach taken by the Tertiary Preparation Program in introducing students to the culture of the university, recognizing the primacy of the context of academic communication and students' social interactions with and within that context, rather than taking a more abstract, skills-based approach to language development. It was appropriate, therefore, that the placement test took a corresponding approach.

Similarly to Hymes, Habermas (1970) recognizes that language use occurs within a communicative context and emphasizes the sociological and functional aspects of language usage. In this, his interest is in a pragmatic approach to language that recognizes that language is used to perform speech acts within a given context and situation with the overall purpose of social cooperation. Again, this formulation is compatible with the learning and teaching approach adopted in TPP: the focus is on developing students' ability to perform tasks (often communicative tasks, such as composing an essay) within the academic context.

Halliday (1978), in pursuing a "psychologically and sociologically realistic overall theory of language and its functions" also argues for the inseparability of language from sociocultural context and further develops the functional approach. Halliday highlights language as the means by which social groups interact and as essential for group identity and belonging. The essential elements of language with which Halliday concerns himself, including the text, the text variety or register, the code, the situation, the linguistic system and the social structure, provide a theoretical justification not only for the functional approach that we adopted in developing the test but are also compatible with contemporary approaches that recognize that learning occurs in an interactive sociocultural environment (Henderson & Cunningham, 1994), and that foreground the importance of a sense of group identity and belonging for the learner (Gopalan & Brady, 2020) – all the more important for students who have faced previous barriers to academic attainment.

On reflection, then, the interrelated theoretical stances of these three theorists provided a firm and appropriate justification for the functional, skills-neutral approach of the ALLTest and were, furthermore, compatible with the contemporary views of academic literacies and

learning in general, cognizant as these views are of the importance of environment, social interaction and group belonging, that were manifested in TPP.

Phase 5: the blueprint revised and finalized: the ALLTest

The development and implementation of the three tiers of the ALLTest occurred over iterative stages that allowed for the selection of the best-performing items from the initial pool of 300 multiple-choice items for Tier 1 and Tier 2 and the determination of cut points for Tiers 1, 2 and 3.

In the first step of the development phase, we tested the multiple-choice items and, by implementing the Tier 3 written test in place of the C-Test that had been the previous testing regime, we determined a cut point that would help us make a recommendation for each applicant to the program as to their enrollment in either of the academic literacy and communication courses (*Communicating at University A* or the more advanced *Communicating at University B*).

The team had devised a surplus of multiple-choice items to allow for the ultimate selection of the specified number of items for Tiers 1 (30 items) and 2 (80 items) based on their level of difficulty, their power to differentiate test-takers, and their internal reliability. Because these items reflected closely the course content of the two courses, and because the test was to be targeted at students of a very similar demographic and level of academic knowledge, the team decided to test the items not as part of a comprehensive pilot test – doing so for 300 items, in any case, would have been too large a burden on a pilot group, and almost certainly would have resulted in a response effect – but by embedding the items in "digestible" chunks in the course content of each of the courses.

Items from the various sub-tests were embedded in appropriate places in the learning progression of each course, and students in these courses completed them voluntarily as supplementary learning material. In all cases, so that the relevant statistical comparisons could be made across items, sub-tests were embedded in their entirety. In the course *Communicating in University A*, for example, the sub-test on scrambled sentences was embedded as supplementary practice in a section of the learning progression where students had completed a lesson on writing cohesive paragraphs. Because the courses are delivered at least partly online, the sub-tests had been transposed into the Moodle quiz module. The module facilitates the administration and automated marking of tests (students were given constructive feedback in the form of learning recommendations based on their scores), but also records students' responses to each item (i.e., *A*, *B*, *C*, *D* or *E*), thus allowing for the individual item analysis that would help the team determine which items to retain, discard, or perhaps amend.

After collecting data over two semesters, we achieved sample sizes of between 70 and 355 on the sub-tests, allowing for the analysis that would assist in selecting items for Tiers 1 and 2. The voluntary nature of the course-embedded sub-tests and factors such as attrition of student numbers across each respective semester account for inconsistencies in sample sizes across sub-tests.

Table 6. Selected examples of items that did not perform satisfactorily in pilot tests.

Sub-test/Item	*Discrimination indices rpBis*	*Item difficulty P-value*	*Internal consistency reliability Alpha*	*Flag(s)*
Scrambled text (*N* = 101)				
2	**0.08**	0.29	**0.47**	K, **LR,**
3	0.41	**0.06**	**0.44**	**LP**
One-word vocabulary (*N* = 116)				
9	**0.13**	**0.96**	0.89	**HP, LR**
10	0.38	**0.94**	0.89	**HP**
Two-word vocabulary (*N* = 93)				
7	–0.15	0.05	0.81	K, **LP, LR**
9	0.86	0.40	0.80	**HP**
Interpreting graphic and visual information (*N* = 70)				
1	**0.07**	**0.03**	0.86	**LP, LR**
20	0.16	0.24	0.86	K
Register and text type (*N* = 355)				
4	**0.12**	**0.98**	0.76	**HP, LR**
5	0.42	**0.88**	0.73	**HP**
Text comprehension (*N* = 139)				
34	0.73	**0.85**	0.96	**HP**
81	**0.05**	0.32	0.96	K, **LR**
Grammar and text relations (*N* = 245)				
3	**0.10**	0.71	0.81	**LR**
4	0.25	**0.93**	0.79	**HP**

Notes. Values not meeting thresholds are in bold text. K = key potentially incorrect; LR = *rpBis* too low; LP = P too low; HP = *P* too high.

Item selection for tiers 1 and 2 of the ALLTest

Item selection was based on classical test theory analysis (see Allen & Yen, 2002), using the Iteman 4.3 application. Three main criteria were used for evaluating and selecting items from each of the pilot sub-tests:

- *The ability of the item* to discriminate levels of ability amongst test-takers, as indicated by the point-biserial correlation (*rpBis*) score. Pilot items that displayed positive *rpBis* scores of 0.2 (acceptable, but needing review; ranging to 1.0, very strong) and above were deemed suitable for selection.
- *The ability of the item to indicate the level of facility of the item* (or how easy or difficult it is), as indicated by the *p*-value. An ideal *P*-value is 0.5 (Guyer & Thompson, 2011); items in the range of 0.15 (very difficult) – 0.84 (very easy) were considered acceptable for selection.
- *The internal consistency reliability of the item*, as indicated by the Alpha-value (Alpha ≥0.8 being the target)

Data from each multiple-choice item were also examined for instances in which the distractors had a higher *rpBis* than the correct (or key) answer; this would indicate that higher scoring test-takers were tending to select an answer other than the key, potentially because the key itself was incorrect (such items are indicated by *K* in Table 6).

Once unacceptable items were eliminated based on these criteria, the best-performing items from each subtest were selected for Tier 2, while the remaining ones were allocated to Tier 1.

Overall, the pilot showed that many of the remaining items were easier (higher *P*-values) than intended when they were devised. As a result, we preferred more difficult items where the discriminatory value was at least satisfactory. Priority in these criteria was given to selecting items for Tier 2 (this tier being the one that would be the most important in determining the placement recommendation for each student). The relative ease of the items, and thus of the overall test, was acceptable given both the adequate power of the selected items to discriminate and the principle of administering a humane test that would not be confidence-eroding for students who have been marginalized in the past.

Additional factors considered in the selection of items were whether they were inadvertent duplications and whether the integrity of question pairs needed to be maintained (particularly in the case of the *Grammar and text relations* sub-tests). As a result, at times the team selected an item of less discriminatory value or greater facility than the one indicated by the statistical analysis.

Implementation of the ALLTest

Once we had selected the best-performing items, we implemented Tiers 1 and 2 by placing them within an online Moodle-based course environment (see Figure 2) which provided commencing students with pre-enrollment advice and placement recommendations.

We developed an automated process for the program's student-facing self-enrollment platform wherein intending TPP students would complete Tier 1—the 30-item practice test – before they could access Tier 2, the main 80-item test which would determine placement

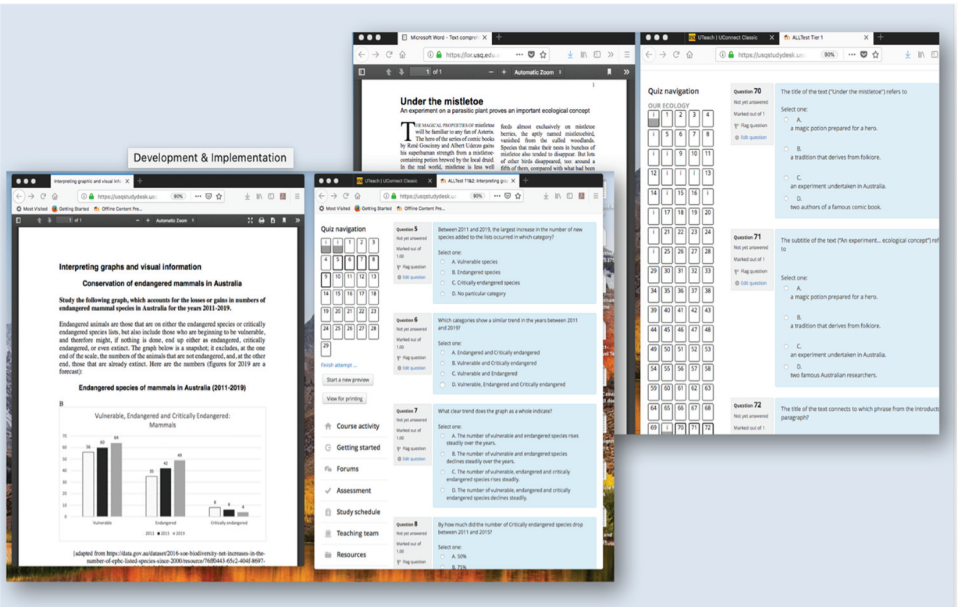

Figure 2. Moodle implementation of the ALLTest: Text comprehension and interpreting graphs and visual information subtests.

recommendations. Students attempted one or both tiers voluntarily (not all who completed Tier 1 elected to attempt Tier 2).

To ensure validity and test security, question items were randomized and each student was allowed one attempt only at each tier. The Moodle quiz module provided a placement recommendation for Tier 2 based on students' scores in relation to the cut point of 68/80.

While Tier 1 also provided general feedback to the students, we decided, after further reflection, to alter the design specifications for this Tier by designating it as a practice test only: the initial design specified that students who scored 80% or above in Tier 1 would be eligible for placement in *Communicating at University B*, thus foregoing Tier 2. We decided, instead, that once we had sufficient data, we would conduct a regression analysis to determine the extent to which a certain score in Tier 1 could predict another (specifically the cut score in Tier 2). This would eliminate arbitrariness in determining a placement score in Tier 1 while retaining its primary function in familiarizing students with the test format and environment.

Both tiers were monitored after their implementation to ensure that there were no technical issues and to re-adjust the cut score in Tier 2, if required, to reflect the 53rd percentile, the determination of which we explain below. During the initial two semesters of the test's implementation, no significant technical problems were detected, while the cut-point was adjusted by only one point to continue to reflect the 53rd percentile.

Determining a cut point for placement levels

To determine a cutoff score for Tier 2, a percentile score was extrapolated from Tier 3, which had, as a result of trials, been calibrated to the C-test. This trial yielded a sample of 869 and was used for placement recommendations in subsequent semesters.

To calibrate Tier 3 to the C-test, each candidate's essay response in the Tier 3 trial was given two scores by the experienced raters: one that was a binary determination of level (*Communicating at University A* or *Communicating at University*

B) based on the rater's knowledge of the respective courses and previous versions of the placement test (i.e., the C-test); and the other was a score derived from the rubric designed for the Tier 3 task. A frequency table generated by a chi-square analysis indicated that, from a maximum of 20 points, most students who scored 12 and above had been placed in the more advanced course.

The cut-point of 12 was confirmed by examination of a receiver operating characteristic (ROC) curve – an analytic procedure adopted from medical diagnosis that is used to determine the probability of a relationship between a certain diagnostic score and a known underlying medical condition (for the case of our analysis, the "underlying condition" was deemed to be the candidate's academic literacy development at the more advanced level). By applying the cut point of 12 and above to all collected test scores, we discovered that 53% of test-takers received a recommendation for advanced placement.

Once test items had been selected for Tier 2, the 53rd percentile was applied to the range of scores in each item to determine the cut-point for each. From this, we calculated an overall cut point of 67.56 out of 80 (or 84.45%). Candidates who scored 68 or above (after rounding) would be given a recommended placement of *Communicating at University B*. This was regarded as a preliminary cut point.

Performance of tiers 1 and 2

Once we had gathered an adequate sample ($N = 806$ for Tier 1; $N = 681$ for Tier 2), we used classical test theory and Rasch analysis (Rasch, 1980) to evaluate the 30 items of Tier 1 and the 80 items of Tier 2, respectively. The non-parametric classical test theory analysis (conducted using TiaPlus and Iteman) was consistent with the previous phase of development (item selection), while the much larger sample size in this later stage of analysis allowed us to triangulate and confirm these results through parametric Rasch analysis.

In Tier 1, an Iteman-generated analysis indicated that all 30 items were performing well (see Table 7), with very good overall reliability (a = 0.79) for a 30-item test (40 items are usually recommended to obtain comparable levels of reliability).

In Tier 2, the analyzes revealed that, with one exception, all 80 items were performing well (see Table 8), with an excellent overall reliability score (a = 0.95).

Discussion

Key imperatives underpinning the development of the ALLTest included the need for a more valid and reliable placement test that would ensure students could be provided with individualized and accurate enrollment advice, along with the need for an operationally more efficient assessment process that could be sustained as enrollment numbers continue to expand. This latter aim was an internal requirement rather than a pedagogically informed ambition related to the most supportive student journey. The test's implementation has also positively impacted student empowerment levels and could inform nuanced and evidence-based curriculum reform.

Firstly, during the development phase, project members were concerned that test questions could prove too challenging for students who were new to academic literacy, may have minimal digital literacy skills, and may be discouraged by the time and attention to detail the test required. However, these concerns proved unfounded. After initial trialing, Tier 2 of the test is now undertaken by approximately 65% of all students applying for the course (approximately 70% complete Tier 1, the practice test), with the remaining students opting to enroll in the introductory academic literacy course as a preference. This suggests students find the test easily accessible, engaging, and of value in the enrollment process, while also acknowledging that students have adequate intrinsic motivation to complete the tests as an entry point to the program. In addition, the online quiz module through which the test is administered allows for a monitored and timed test process, ensuring greater reliability and comparability of student results.

Table 7. Summary statistics of ALLTest Tier 1.

Score	Items	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis	Alpha
Scored Items	30	21.31	4.83	5	30	0.71	0.29	0.79
Vocabulary (one-word)	6	5.01	1.00	1	6	0.83	0.13	0.29
Vocabulary (two-word)	4	3.42	0.79	0	4	0.85	0.21	0.33
Scrambled text 1	4	2.70	1.25	0	4	0.68	0.21	0.67
Scrambled text 2	4	2.28	1.67	0	4	0.57	0.30	0.87
Grammar & text relations	12	7.90	3.27	0	12	0.66	0.41	0.83

Note. $N = 806$; Livingston index of classification consistency at cut-score of 15 was 0.92.

Table 8. Summary statistics of ALLTest Tier 2.

Score	Items	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis	Alpha
Scored Items	80	58.71	15.50	7	80	0.73	0.43	0.95
Vocabulary (one-word)	8	6.47	1.69	1	8	0.81	0.35	0.66
Vocabulary (two-word)	6	4.47	1.38	0	6	0.75	0.32	0.50
Scrambled text	5	4.05	1.67	0	5	0.81	0.44	0.91
Register & text type	5	4.24	1.09	0	5	0.85	0.27	0.60
Graphic & visual information	10	6.86	2.57	0	10	0.69	0.40	0.78
Text comprehension	30	21.93	7.11	0	30	0.73	0.48	0.92
Grammar & text relations	16	10.69	4.80	0	16	0.67	0.49	0.92

Note. $N = 681$; Livingston index of classification consistency at cut-score of 40 was 0.98.

The ALLTest has achieved, moreover, a significant reduction in the administrative burden of enrolling students. Students are now able to access and complete the test independently through an online platform that articulates clear instructions for students. For Tier 1 and Tier 2, results are automatically generated and are provided to students immediately, allowing them to make informed self-enrollment decisions. Tier 3, undertaken primarily by incarcerated students without access to an online platform, is now the only tier that requires manual marking by academic staff.

While the ambition of streamlining administrative processes was always subordinate to the more significant work of appropriately testing and placing students, existing processes had also tended to subordinate and disempower students. Many were bulk enrolled in the lower-level courses and had limited agency in choosing their starting level. As noted above, the high number of applicants completing the ALLTest not only indicates that the test structure and volume are neither overwhelming nor off-putting, but it also indicates that students are choosing to engage in decision-making about their course selection. The students making these decisions come from cohorts comprising people with low self-confidence and a limited family history of higher education participation.

As Klinger and Tranter (2009) point out, entering university even via the supportive pathway of an Enabling program involves some degree of cultural displacement and disorientation. We propose that the design and implementation of the ALLTest is a further important strategy to build confidence. As outlined above, the test structure and content demystifies academic literacy by breaking down common academic tasks into easily recognized subconstructs while gently introducing notions of key competencies (O'Rourke et al., 2019).

Limitations

We acknowledge that some of the limitations identified in the literature, such as those relating to weak predictability and narrow focus, may not be surmountable with a one-time, limited-scope test. Our approach, however, has been to mitigate some of these challenges by ensuring that the ALLTest is closely aligned with functional academic literacy, and that it is used not as a gatekeeping measure but for informed self-enrollment based on a valid and reliable diagnostic instrument.

While a one-time test remains less than ideal, many of our non-traditional students do not have recent (or any) Year 12 results that could provide a viable placement alternative. Over time, however, we hope that the ALLTest (or its successor) may be a part of a more

holistic approach to assessment that may include other methods, such as student portfolios. These alternative or complementary methods provide us with a direction for further research.

We should also note that the scope of this paper concerns itself with the development and initial implementation of the ALLTest; while valid questions remain as to its potential to predict literacy development, academic success, and retention, these are more appropriately left to a further stage of maturity once sufficient *in situ* data have been collected.

An additional limitation is that relating to the generalizability of the test. As outlined in this paper, we elected to develop a contextualized test because such a measure would have direct relevance and applicability to our students at UniSQ and would align with the specific functional academic literacy knowledge embedded in our courses. We recognize, too, that language is used within a specific sociocultural context – indeed, this was one of our reasons for avoiding a conventional skills-neutral approach. We acknowledge, however, that a highly contextualized test has limited generalizability beyond our specific context. For this reason, we have described in detail the Australian higher education, widening participation, and institutional context: this detail may be of some use to readers who are able to identify analogous conditions in their own context.

With these limitations to the generalizability of the actual test *product* in mind, however, we believe that the *process* adopted through the five-phase applied linguistics model is one that can be widely adopted by others who seek to develop their own contextualized academic literacy placement test, and we hope that the phases that we have described in detail will aid them in doing so.

Conclusion

In this article, we have presented a case study of the use of a five-phase model of applied linguistic design to create the Academic Literacy Level Test (ALLTest). The context for the case study was an Enabling program in Australia intended to facilitate a pathway for non-traditional students to enter degree-level study. The intention of the test creation was a reliable placement test, facilitating responsive pre-entry advice to students on an appropriate entry level to their study of academic literacy and communication. The resulting three-tiered ALLTest provides evidence that the use of the five-phased applied linguistic design – which has at its core the precedence of pragmatic and technical imagination – is a fitting design model for providing a context-sensitive test of academic literacy that, for the institution, is operationally sound, comprehensive and valid, and, for incoming students, provides appropriate and timely advice to empower their decision-making regarding their entry options.

The ALLTest is now a business-as-usual component of the UniSQ TPP. Since its implementation, very few students have queried their placements, indicating strong acceptance of the test advice among students. Further, when a student does query the placement advice, whether because of technical glitches, misunderstandings, or a mismatch of expectations, the Tier 3 test is readily available as a basis for review. While this suggests students are following the ALLTest advice, follow-up research is planned regarding the impact of the ALLTest advice upon student enrollment choices. As thousands of students choose to participate in the testing, it is also generating a substantial body of data, supplemented by other data sources including the progression and retention of the students into a degree. It is

a further benefit of the ALLTest that it will make possible informed decision-making for program development and the suitability of the courses for student cohorts, including the continuity of progression between the two stages of academic communication that currently exist in the program.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Jonathan H. Green http://orcid.org/0000-0003-1468-1970
Charmaine Davis http://orcid.org/0000-0003-2194-2307
Marcus Harmes http://orcid.org/0000-0002-5104-1967
Kate Judith http://orcid.org/0000-0002-2021-9830
Albert Weideman http://orcid.org/0000-0002-9444-634X

References

Agosti, C. I., & Bernat, E. (2018). University pathway programs: Types, origins, aims and defining traits. In C. I. E. Bernat (Ed.), *University pathway programs: Local responses within a growing global trend* (pp. 3–25). Springer.

Allensworth, E. M., & Clark, K. (2020). High school GPAs and ACT scores as predictors of college completion: Examining assumptions about consistency across high schools. *Educational Researcher*, *49*(3), 198–211. https://doi.org/10.3102/0013189X20902110

Allen, M. J., & Yen, W. W. (2002). *Introduction to measurement theory*. Waveland Press.

Armstrong, W. B. (2000). The association among student success in courses, placement test scores, student background data, and instructor grading practices. *Community College Journal of Research & Practice*, *24*(8), 681–695. https://doi.org/10.1080/10668920050140837

Australian Government. (2017). *Higher Education Support Act 2003*.

Baker, S. & Irwin, E. (2012). *A national audit of academic literacies provision in enabling courses in Australian higher education (HE): Report compiled for the Association of academic Language & Learning*. https://www.newcastle.edu.au/__data/assets/pdf_file/0009/349434/Baker-Irwin-2015-AALL-report-final.pdf

Baker, S., & Irwin, E. (2015a). *Enabling Typology*. (https://enablingeducators.org/wp-content/uploads/2020/12/Enabling-Programs-Australia-Typology-KS.pdf

Baker, S., & Irwin, E. (2015b). *A national audit of academic literacies provision in enabling courses in Australian Higher Education (HE)*. http://www.aall.org.au/sites/default/files/Baker-IrwinAALL-rep

Barnett, E., Bergman, P., Kopko, E. M., Reddy, V. T., Belfield, C., & Roy, S. (2018). *Multiple measures placement using data analytics: An implementation and early impacts report*.

Beleche, T., Fairris, D., & Marks, M. (2012). Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. *Economics of Education Review*, *31*(5), 709–719. https://doi.org/10.1016/j.econedurev.2012.05.001

Bennett, A., Motta, S., Hamilton, E., Burgess, C., Relf, B., Gray, K., & Albright, J. (2018). *Enabling pedagogies: A participatory conceptual mapping of practices at the university of newcastle, Australia*. (https://www.newcastle.edu.au/__data/assets/pdf_file/0005/462272/Enabling-Pedagogies-Research-Report.pdf

Brett, T., Trinidad, S., Devlin, M., Harvey, A., Pitman, M., & McKay, J. (2016). Pathways to higher education: The efficacy of enabling and sub-bachelor pathways for disadvantaged students. *Public Library of Science ONE*, *11*(4). https://doi.org/10.1371/journal.pone.0153625

Burdman, P. (2012). Where to begin? The evolving role of placement exams for students starting college. *Jobs for the Future*. www.jff.org/idea/where-begin-evolving-role-placement-exams-students-starting-college/

Burke, P. J. (2013). *The right to higher education: Beyond widening participation*. Routledge.

Cazden, C. B. (2011). Dell hymes's construct of "communicative competence". *Anthropology & Education Quarterly*, *42*(4), 364–369. https://doi.org/10.1111/j.1548-1492.2011.01144.x

Chesters, J., & Watson, L. (2016). Staying power: The effect of pathway into university on student achievement and attrition. *Australian Journal of Adult Learning*, *56*(2), 225–249.

Chomsky, N. (1965). *Aspects of the theory of syntax*. M.I.T. Press.

Christie, H., Munro, M., & Fisher, T. (2004). Leaving university early: Exploring the differences between continuing and non-continuing students. *Studies in Higher Education*, *29*(5), 617–636. https://doi.org/10.1080/0307507042000261580

Cliff, A., & Hanslo, M. (2010). The design and use of "alternate" assessments of academic literacy as selection mechanisms in higher education. *Southern African Linguistics and Applied Language Studies*, *27*(3), 265–276. https://doi.org/10.2989/SALALS.2009.27.3.5.939

Cooper, E. E., McGee, J. R., Levine-Brown, P., & Bolt, L. (2019). The effectiveness of redesigns in developmental education. *Journal of Developmental Education*, *43*(1), 20–25,32. Retrieved from https://www.proquest.com/scholarly-journals/effectiveness-redesigns-developmental-education/docview/2317570374/se-2

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238. https://doi.org/10.2307/3587951

Davis, C., Syme, S., Cook, C., Dempster, S., Duffy, L., Hattam, S., Lambrinidis, G., Lawson, K., & Levy, S. (2023). *Report on Bechmarking of Enabling Programs across Australia to the National Association of Enabling Educators of Australia (NAEEA)*. https://enablingeducators.org/reports/

Department of Education Skills and Employment. (2020). *uCube higher education statistics*.

Dimova, S., Yan, X., & Ginther, A. (2020). *Local language testing: Design, implementation, and development*. Routledge.

Elder, C., & Read, J. (2015). Post-entry assessments in other countries. In *Assessing English proficiency for university study* (pp. 70–92). Palgrave Macmillan UK. https://doi.org/10.1057/9781137315694_4 .

Elliot, N., Deess, P., Rudniy, A., & Joshi, K. (2012). Placement of students into first-year writing courses. *Research in the Teaching of English*, *46*(3), 285–313. https://doi.org/10.58680/rte201218457

Finnegan, F., & Merrill, B. (2017). "We're as good as anybody else": A comparative study of working-class university students' experiences in England and Ireland. *British Journal of Sociology of Education*, *38*(3), 307–324. https://doi.org/10.1080/01425692.2015.1081054

Fulcher, G. (1997). An English language placement test: Issues in reliability and validity. *Language Testing*, *14*(2), 113–139. https://doi.org/10.1177/026553229701400201

Golder, K. J. (2006). *The usefulness and ethical status of language entrance/placement testing at a post-secondary institution*. University of British Columbia.

Gopalan, M., & Brady, S. T. (2020). College students' sense of belonging: A national perspective. *Educational Researcher*, *49*(2), 134–137. https://doi.org/10.3102/0013189x19897622

Gordon, R. (2006). *Why do we do what we do, and how do we know if it is working?*. ACCUPLACER® Placement Conference, Washington, DC.

Grebennikov, L., & Shah, M. (2012). Investigating attrition trends in order to improve student retention. *Quality Assurance in Education*, *20*(3), 223–236. https://doi.org/10.1108/09684881211240295

Guyer, R., & Thompson, N. A. (2011). *User's manual for xcalibre item response theory calibration software, version 4.1.3*. Assessment Systems Corporation.

Habermas, J. (1970). Towards a theory of communicative competence. In H. P. Dreitzel (Ed.), *Recent sociology* (Vol. 2, pp. 115–148). Collier-Macmillan.

Halliday, M. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.

Henderson, R. W., & Cunningham, L. (1994). Creating interactive sociocultural environments for self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self regulation of learning and performance* (pp. 255–281). Routledge.

Hodges, B., Bedford, T., Hartley, J., Klinger, C., Murray, N., O'Rourke, J., & Schofield, N. (2013). *Enabling retention: Processes and strategies for improving student retention in university-based enabling programs. Final report 2013*. Australian Government Office for Learning and Teaching.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: selected readings* (pp. 269–293). Penguin.

Kirst, M. W. (1998). *Improving and aligning K-16 standards, admissions, and freshman placement policies*. O. o. E. R. a. I. (ED). https://eric.ed.gov/?id=ED434608

Klinger, C., & Murray, N. (2012). Tensions in higher education: Widening participation, student diversity and the challenge of academic language/literacy. *Widening Participation and Lifelong Learning*, *14*(1), 27–44. https://doi.org/10.5456/WPLL.14.1.27

Klinger, C., & Tranter, D. (2009). *Firm foundations for the future*. Third National Conference of Enalbing Educators of Australia, Toowoomba, Australia.

Krashen, S. D. (1987). *Principles and practice in second language acquisition*. Prentice Hall International.

Lea, M. R. (2017). Academic literacies in theory and practice. In B. V. Street & S. May (Eds.), *Literacies and language education* (pp. 147–158). Springer International Publishing. https://doi.org/10.1007/978-3-319-02252-9_19.

Lisciandro, J., & Gibbs, G. (2016). OnTrack to university: Understanding mechanisms of student retention in an Australian pre-university enabling program. *Australian Journal of Adult Learning*, *56*(2), 198–224.

Lomax-Smith, J., Watson, L., & Webster, B. (2011). *Higher education base funding review: Final report*.

McKay, J., Pitman, T., Devlin, M., Trinidad, S., Harvey, A., & Brett, M. (2018). The use of enabling programs as a pathway to higher education by disadvantaged students in Australia. In C.-I.-B.-A. Eva (Ed.), *University pathway programs: local responses within a growing global trend* (pp. 45–66). Springer International Publishing.

Mokher, C., McCoy, K., Henning, H., Fluker, C., & Park-Gaghan, T. (2023). Placing students for success: Florida's statewide reform of alternative methods for college course placement. *Community College Journal of Research and Practice*, *48*(3), 1–18. https://doi.org/10.1080/10668926.2023.2189186

Morante, E. (1989). Selecting tests and placing students. *Journal of Developmental Education*, *13*(2), 2–4. 6.

Murphy, S., & Yancey, K. B. (2009). Construct and consequence: Validity in writing assessment. In C. Bazerman (Ed.), *Handbook of research on writing: History, school, individual, text* (pp. 448–473). Lawrence Erlbaum Associates.

National Association of Enabling Educators of Australia. (2019). *Common learning outcomes for enabling courses in Australia*. https://enablingeducators.org/wp-content/uploads/2019/11/2019-Learning-Outcomes_Enabling-Courses_Australia_AB-002.pdf

Neghina, C. (2015). *New routes to higher education: The global rise of foundation programmes study portals*. Cambridge English.

Noble, J. P., Schiel, J. L., & Sawyer, R. L. (2003). Assessment and college course placement: Matching students with appropriate instruction. In J. E. Wall & G. R. Walz (Eds.), *Measuring up: Assessment issues for teachers, counsellors and administrators* (pp. 297–311). Office of Educational Research and Improvement (ED).

O'Rourke, J. A., Relf, B., Crawford, N., & Sharp, S. (2019). Are we all on course? A curriculum mapping comparison of three Australian university open-access enabling programs. *Australian Journal of Adult Learning*, *59*(1), 7–26.

Patterson, R., & Weideman, A. (2013). The refinement of a construct for tests of academic literacy. *Journal for Language Teaching*, *47*(1), 125–151. https://doi.org/10.4314/jlt.v47i1.6

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* The University of Chicago Press.

Scott-Clayton, J., Crosta, P. M., & Belfield, C. R. (2014). Improving the targeting of treatment: Evidence from college remediation. *Educational Evaluation and Policy Analysis*, *36*(3), 371–393. https://doi.org/10.3102/0162373713517935

Sebolai, K. (2016). Distinguishing between English proficiency and academic literacy in english. *Language Matters*, *47*(1), 45–60. https://doi.org/10.1080/10228195.2015.1124281

University of Southern Queensland. (2022). *The University of Southern Queensland Annual Report 2021*. Office of the Vice-Chancellor.

Valentine, J. C., Konstantopoulos, S., & Goldrick-Rab, S. (2017). What happens to students placed into developmental education? A meta-analysis of regression discontinuity studies. *Review of Educational Research*, *87*(4), 806–833. https://doi.org/10.3102/0034654317709237

Venezia, A., Bracco, K. R., & Nodine, T. (2010). *One-shot deal? students' perceptions of assessment and course placement in california's community colleges*. WestEd.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Weideman, A. (2019a). Definition and design: Aligning language interventions in education. *Stellenbosch Papers in Linguistics Plus*, *56*(1), 31–46. https://doi.org/10.5842/56-0-782

Weideman, A. (2019b). Degrees of adequacy: The disclosure of levels of validity in language assessment. *Koers*, *84*(1), 1–15. https://doi.org/10.19108/KOERS.84.1.2451

Weideman, A. (2019c). Validation and the further disclosures of language test design. *Koers*, *84*(1), 1–10. https://doi.org/10.19108/KOERS.84.1.2452

Weideman, A. (2021). A skills-neutral approach to academic literacy assessment. In A. Weideman, J. Read, & T. du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and Transformation* (pp. 22–51). Multilingual Matters. https://doi.org/10.21832/9781788926218.

Weideman, A., & Deygers, B. (2024). Validity and validation: An alternative perspective chapter 2. In M. R. Salaberry, A. Weideman, & W.-L. Hsu (Eds.), *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 28–50). Routledge. https://doi.org/10.4324/9781003384922-3.

Weideman, A., Read, J., & Du Plessis, T. (2020). *Assessing academic literacy in a multilingual society: Transition and transformation* (Vol. 84). Multilingual Matters.

Willans, J., & Seary, K. (2018). Why did we lose them and what could we have done? *Student Success*, *9*(1), 47–60. https://doi.org/10.5204/ssj.v9i1.432