
2018

The National Benchmark Quantitative Literacy Test for Applicants to South African Higher Education

Vera Frith

University of Cape Town, vera.frith@uct.ac.za

Robert N. Prince

University of Cape Town, robert.prince@uct.ac.za

Follow this and additional works at: <https://digitalcommons.usf.edu/numeracy>

Recommended Citation

Frith, Vera, and Robert N. Prince. "The National Benchmark Quantitative Literacy Test for Applicants to South African Higher Education." *Numeracy* 11, Iss. 2 (2018): Article 3. DOI: <https://doi.org/10.5038/1936-4660.11.2.3>

Authors retain copyright of their material under a [Creative Commons Non-Commercial Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

The National Benchmark Quantitative Literacy Test for Applicants to South African Higher Education

Abstract

The National Benchmark Test Project (NBTP) was commissioned by Higher Education South Africa in 2005 to assess the academic proficiency of prospective students. The competencies assessed include quantitative literacy using the NBTP QL test. This instrument is a criterion-referenced multiple-choice test developed collaboratively by South African academics and provides complementary information to that provided by the norm-referenced school-leaving examination. In this paper we outline the theoretical framework that provides the foundation for the NBTP QL test and describe the test construct. In the QL test specifications, there are three dimensions specified for each item: the competencies (reasoning and behaviour) that are required to answer the item correctly, the main mathematical and statistical ideas the item addresses, and a characterisation of the level of cognitive processing the item calls for. The results are reported using benchmarks which place students' scores into proficiency bands which indicate the extent to which curricula should be responsive to their level of preparedness. We discuss the extent to which the NBTP QL test is succeeding in contributing to meeting the goals of the NBT project. The test is intended to provide institutions with information that will assist with admissions and placement decisions, but the QL test scores are not used uniformly for these purposes across the higher education sector. The NBTP QL test results show that the majority of students are severely underprepared for the QL demands of higher education and that a comprehensive systemic response requiring curriculum change at many levels is required.

Keywords

quantitative literacy, numeracy, mathematical literacy, assessment, testing

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

Cover Page Footnote

Vera Frith is the coordinator of the Numeracy Centre, a unit within the Centre for Higher Education Development at the University of Cape Town. Her primary interests are the quantitative literacy development of university students and the appropriate curriculum for this purpose.

Robert Prince is the test development coordinator in the Centre for Educational Testing for Access and Placement at the University of Cape Town. His interests include the assessment of academic literacies, including quantitative literacy, and the manner in which this assessment can inform the development of appropriate curriculum interventions.

Introduction

Most students leaving the school (K-12) system in South Africa are, to varying extents, underprepared for the demands of higher education, and the higher education system is underprepared for many of the students they do admit. Since 1994 the number of students in higher education has almost doubled, but the curriculum has not changed in a way that takes into account the changing nature of the student body. This curricular inertia results in only 27% of students graduating in minimum time and 55% unlikely ever to graduate (Scott et al. 2007).¹ The experience of higher education and outcomes for students are highly unequal, as both the school and higher education sectors still labour under the after-effects of the unequal apartheid education system. According to Scott et al. (2007, 42) "... the educational factor to which poor performance is perhaps most commonly ascribed across the higher education sector is student under-preparedness for standard undergraduate programmes." Difficulties experienced in the area of academic literacies, including quantitative literacy (QL), are recognised as major contributors to this problem.

The importance of QL for higher education is widely recognised (see, for example, Steen 2004), and there is also an increasing awareness that many academic disciplines make complex quantitative demands that are often very different from those that are the focus of traditional mathematics courses. Some examples of the QL demands of South African higher education curricula in Health Sciences and Engineering are outlined in Frith and Prince (2009), Frith and Gunston (2011) and Prince and Simpson (2016).

Partly in response to the challenge of student under-preparedness, Higher Education South Africa (HESA) commissioned the National Benchmarks Tests Project (NBTP) in 2005 to assess the academic proficiency of prospective students in higher education. The tests assess proficiency in QL as well as academic literacy for all students. Those students entering disciplines that require them to take a mathematics course in their higher education programme, also take a mathematics proficiency test. While the mathematics test is designed to assess students' abilities to do the mathematics they have learned at school and to transfer it to higher education mathematics courses, the QL test is not directly aligned with the school curriculum or mathematics courses in higher education. These tests are designed to provide information complementary to that provided by the National Senior Certificate (the South African school-leaving qualification taken at the end of grade 12), to assist with selection and placement of students into appropriate courses and programmes in higher education, and to provide information to inform curriculum development.

¹ The minimum time required to complete a qualification depends on the discipline. For example, it takes three years to complete a Humanities degree and four years for an Engineering degree.

As stated in Griesel (2006), the goals of the NBTP are as follows:

- To assess entry-level academic and quantitative literacy and mathematics proficiency of students;
- To assess the relationship between entry level proficiencies and school-level exit outcomes;
- To provide a service to higher education institutions requiring additional information in the admission and placement of students; and

To inform the nature of foundation courses and curriculum responsiveness. The competences that are assessed in the NBTP tests (including QL) are regarded by the higher education sector as key areas in which entering students should have minimum levels of proficiency. These tests are criterion-referenced, and minimum benchmark scores for three different proficiency categories are established. In addition, the description of each proficiency category includes recommendations for the kind of educational provision appropriate for students whose scores are in that category.

The construct of the NBTP QL test is outlined in Frith and Prince (2006) and will be described further in this paper. In practical terms, the test assesses students' abilities to competently interpret and reason with quantitative information presented in a variety of modes in real contexts relevant to quantitative disciplines in higher education. For example, students must apply quantitative procedures in various situations; read and interpret tables, graphs, charts, diagrams, and texts; and integrate information from different sources. The test also assesses the ability to do simple calculations and estimations which may involve multiple steps, to interpret two-dimensional representations of three-dimensional structures, to formulate and apply simple formulae, to reason logically, and to identify trends and patterns in various situations.

In this paper, we will describe the theoretical framework underlying the construct of the NBTP QL test and provide an outline of the test specifications. We will describe the historical development of the test from its antecedents in 1999 and will provide details about aspects of the test production, administration, scoring, and benchmarks setting processes. Finally, we will outline how the test results are used by institutions in South Africa and discuss the extent to which the QL test is succeeding in contributing to the purposes of the National Benchmark Tests Project.

Institutional Context

There are 26 public universities in South Africa. Twelve of them are traditional universities offering theoretically-oriented university degrees; eight are universities of technology offering vocationally-oriented diplomas and degrees; and six are comprehensive universities offering a combination of both types of qualification. In addition, there are several private universities and colleges offering tertiary

qualifications. There are three different levels of tertiary undergraduate qualification: degrees, diplomas, and higher certificates, each of which have different entrance requirements. Different levels of performance on the national examination taken at the end of grade 12, the National Senior Certificate (NSC), have to be achieved in order to gain access to these programmes, with the most stringent being the entrance to degree programmes.

The public residential universities are, in general, large, having an average total undergraduate and postgraduate enrolment of approximately 25,000 (minimum 6,700 and maximum 60,000). But by far the largest enrolment (300,000) is in the University of South Africa, which is a comprehensive university providing distance education. The public universities together have about 1 million students, but this number reflects a participation rate of less than 20% of the population aged 20-24 years (Council on Higher Education 2017).

All of South Africa's public universities are members of the Universities South Africa, formerly known as Higher Education South Africa (HESA), a non-profit, non-statutory association which aims to promote a more inclusive, responsive, and equitable national system of higher education.

Our institution, the University of Cape Town (UCT) is a traditional residential university with approximately 20,000 undergraduate students and 10,000 postgraduates. The university is divided into relatively autonomous discipline-based divisions (faculties), each headed by a Dean. These are the Faculties of Humanities, Science, Engineering and the Built Environment, Commerce, Law, Health Sciences, as well as the Centre for Higher Education Development (CHED), which operates as a seventh faculty. The units to which the authors belong are situated in CHED. One of the largest departments in CHED is the Academic Development Programme, which is responsible for extended and augmented educational provision for students who require it in all the faculties.

The Antecedents of the NBTP QL Test at UCT

The Numeracy Centre (NC) at UCT was established in 1999 in response to the stipulations in the UCT Strategic Planning Framework 1997 – 2000 document that all graduates be effectively literate, effectively numerate, and computer literate at the appropriate level. The Numeracy Competency Test project was established at the same time as the NC and formed an important part of its brief. Its initial goal was to develop a measure of the levels of numeracy (QL) of first year Humanities students at UCT. Such a test would allow staff in the NC to give a reasoned response to the question of whether an add-on numeracy course should be mounted for those whose numeracy competency was demonstrably too low.

We will briefly describe the Numeracy Competency Test project and how it developed because it was in effect the forerunner of the NBTP QL test. It was by

working on this project (and by teaching a numeracy course) that we and other NC staff developed our views on the theoretical foundations of QL and also developed the necessary abilities to take responsibility for the development of the NBTP QL test when that project came into being.

Because in 1999, we (the authors) were newly appointed to the NC and came from mathematics education backgrounds, we had much to learn about numeracy. To guide us we adopted the following definition from the International Life Skills Survey (Gal et al. 2005):

Numeracy involves abilities, behaviours, and processes which include identifying, interpreting, acting upon and communicating about mathematical information flexibly in real contexts and situations, to enable full, critical, and effective participation in a wide range of life roles.

An iterative model for the development of the test was adopted as follows: In 1999, a pilot test covering a limited range of numeracy topics was administered to a small group of volunteer Humanities students. In February 2000, a revision of this test was piloted on the entire Humanities intake.

The aim of the pilot was to pre-test different types of questions and to obtain information about the “level of numeracy” of students in Humanities in order to inform the design of the test. For the pilot test, we addressed a subset of topics which dealt with fractions, decimals, percentages, graphs and tables, and a few examples of “every day” mathematics using examples that would not require the use of a calculator. The insights gained from running these pilot tests was extremely useful in providing information about the level of difficulty that would be appropriate in the actual test.

As far as the students were concerned, it was termed “The Numeracy Questionnaire” in order to reduce anxiety on their part and to reinforce that the results would be used to develop interventions to assist them in their studies and that they would suffer no negative consequences of performing poorly or “failing.”

Although the initial stage of the project was focused on measuring the competency of first-year Humanities students, the ultimate goal of the project was to develop a test that could be used to ascertain whether students across all disciplines could be identified as “appropriately numerate” before embarking on their studies, and eventually whether on graduating they could be assessed to be numerate as graduates. The test was also intended to provide information about students’ competencies to inform the process of course and curriculum design.

From 2001, a new version of the questionnaire was developed annually and used to assess the numeracy competency of incoming students at UCT. At first it was used only by Humanities to inform curriculum development in the first course run by the NC. Later this assessment was used to identify students needing to do

the QL course as part of their humanities foundation programme for quantitative disciplines (such as economics and psychology). The test was also used in Law to select students for the QL course provided by the NC specifically for law students. The medical students also completed the questionnaire and the results were instrumental in the establishment of the NC's QL interventions for first-year medical students.

When the NBT project was introduced by HESA, the QLT had an established role at UCT, and we possessed theoretical understanding and expertise necessary for developing QL tests for incoming students. At the time that the NBT project was conceptualised, the question was raised whether QL should be tested as part of the Academic Literacy or Mathematics Constructs, or whether there should be a separate assessment for QL based on an independent QL construct. The experience gained from running the Numeracy Competency Test project allowed us to argue successfully for the establishment of QL as a separate domain and provided insights into how the NBTP QL test result might be used in higher education generally.

When HESA looked for expertise for hosting the NBT project, the experience of the NC in QL testing contributed to the decision to locate the project at UCT. This choice resulted in the then NC Co-ordinator (RP) being identified as the person responsible for the development of the NBTP QL test construct and the tests themselves. In this he was assisted by NC staff, and in particular by VF.

Theoretical Framework and Construct

One of our first tasks for the NBT project was to contribute the chapter explaining the QL test domain in the report "Access and entry level benchmarks: The national benchmark test project" (Griesel 2006) published by HESA in 2006. In this section, we will describe the theoretical framework we presented in that publication, which underpinned the test construct that we created at that time and which has been used since then.

In order to produce a definition of QL that could be used as a basis for the test construct, we drew on the work of authors such as Chapman and Lee (1990), Chapman (1998), Steen (2001), Baynham and Baker (2002), Jablonka (2003), and Street (2005). We were strongly influenced by the assessment frameworks of the Adult Literacy and Lifeskills, ALL Survey (Gal et al. 2005), NAEP (National Assessment Governing Board 2004), Programme for International Student Assessment, PISA (OECD 2003) and the Third International Mathematics and Science Study, TIMSS (Mullis et al. 2003). Another important contribution was our experience of running the Numeracy Competency Test project of the NC at UCT. (See Frith et al. 2003 and Prince et al. 2004.)

At the time, there was an ongoing debate about the meaning of the term "Quantitative Literacy" or "Numeracy" and its relationship to "literacy" and to

“mathematics.” This debate was exemplified by the various articles in *Mathematics and Democracy: The Case for Quantitative Literacy* (Steen 2001), which strongly influenced our thinking. However, we also believed (Archer et al. 2002) and still believe (see, for example, Frith and Lloyd 2016 and Prince and Simpson 2016) that it is useful and generative to conceptualise QL as social practice in line with the approach of the New Literacy Studies (Street 2005; Street and Baker 2006; Kelly et al. 2007), which is influential in British and Australian literacy and numeracy circles. It is clear that QL cannot be seen as an identifiable set of skills that can be taught and learned without reference to the social contexts or the academic discourses where they are applied. Given that in higher education there are many different disciplinary discourses requiring different types of literacy, there will also be different QL practices associated with different academic disciplines. (For examples, see Prince and Archer 2006, Prince and Archer 2008, Frith and Prince 2009, Frith and Gunston 2011, and Prince and Simpson 2016). This multi-faceted nature presents a challenge for designing a general purpose QL test for all higher education entrants.

In developing a test construct, we took the view that QL can be described in terms of 1) the contexts in which QL is practiced; 2) the mathematical and statistical content that is required in this practice; and 3) the reasoning and behaviours that are integral to QL practice.

We adopted the following definition of QL, in which all three approaches to the description (contexts, content, and reasoning/behaviours) are embedded:

Quantitative literacy is the ability to manage situations or solve problems in practice, and involves responding to quantitative (mathematical and statistical) information that may be presented verbally, graphically, in tabular or symbolic form. It requires the activation of a range of enabling knowledge, behaviours and processes and it can be observed when it is expressed in the form of a communication, in written, oral or visual mode. (Frith and Prince 2006, 30)

In almost all of the questions in the NBTP QL test, test-takers have to apply quantitative methods and reasoning within a realistic (mostly relevant to higher education) context. The test construct does not specify the contexts for the items, but rather that there is a range of different kinds of contexts, in order that the test has both face validity and relevance for all disciplines in higher education. The challenge is to find contexts that are sufficiently relevant that they motivate test-takers to truly display their potential for quantitatively literate practice.

With regard to content, there can be no exhaustive list of mathematical and statistical content knowledge that is essential for quantitatively literate practice in higher education because this content knowledge will be different for different

academic disciplines. Both Steen (2001) and Hughes-Hallett (2001) made the point that statistics and data handling (rather than traditional school mathematics topics) are very important for QL, and this is certainly true for many academic disciplines. In designing the test construct, we used our experience of teaching QL to students in various disciplines in order to develop a list of the main mathematical and statistical ideas necessary for QL practices in higher education. The emphasis on the application of mathematical and statistical knowledge in context implicitly assumes the importance of the associated quantitative thinking and reasoning. In the literature, we could find no clear definition or characterisation of these “mathematical actions” which include such activities as drawing connections, visualising, questioning, representing, concluding, and communicating (Boaler 2002). However, we recognised that some form of characterisation of these critical competencies must be part of a test construct for QL.

In the QL test specifications, there are three dimensions specified for each item: the competencies (reasoning and behaviour) that are required to answer the item correctly, the main mathematical and statistical ideas the item addresses, and a characterisation of the level of cognitive processing the item calls for. This last was informed by the characterisation used in assessment in the school curriculum (Department of Basic Education, South Africa 2005). The specifications in terms of these three dimensions are shown in detail in Tables 1, 2 and 3 respectively. Items may simultaneously assess more than one competency or mathematical idea, and classifying items according to these two dimensions is relatively uncontroversial. Specifying which cognitive processing dimension an item belongs to is the most difficult and subject to interpretation.

The competence areas and mathematical and statistical ideas dimensions of the specifications allow the test itself as well as individual student results and cohort results to be analysed using clustering of items according to which competence area or mathematical ideas they address. For example, the test can be analysed from the point of view of which competencies (such as reasoning) are required or which quantitative content areas are addressed (such as Quantity; Number and operations; or Relationships, pattern, and permutations). Examples of this kind of analysis are presented in Frith and Prince (2016).

Table 1
Specification for the Quantitative Literacy Test: Classification Based on Competencies

Competence area		Description/specifications	% of test
Comprehending: identifying or locating	Vocabulary	The ability to understand the meanings of commonly encountered “quantitative” terms and phrases (such as percentage increase, rate, approximately, representative sample, compound interest, average, order, rank, category, expression, equation) and the mathematical and statistical concepts (including basic descriptive statistics) that these words refer to. This includes knowledge of systems of units of measurement	15-20
	Representations of numbers and operations	The ability to understand the conventions for the representation of (whole numbers, fractions, decimals, percentages, ratios, scientific notation, measurements, variables) and simple operations (+, -, ×, ÷, positive exponentiation, square roots) on them.	5-10
	Conventions for visual representations	The ability to understand the conventions for the representation of data in tables (several rows and columns and with data of different types combined), charts (pie, bar, compound bar, stacked bar, “broken” line, scatter plots), graphs and diagrams (such as tree diagrams, scale and perspective drawings, and other visual representations of spatial entities).	20 - 25
Acting, interpreting, and communicating	Using representations of data	The ability to derive and use information from representations of contextualised data and to make meaning from this information. For example: Reading values off a chart or observing trends or relationships in tabulated data, using observations of the slope of a graph to derive information about rates, reading off maximum and minimum values.	20 - 25
	Computing	The ability to identify the necessary simple calculations required by a problem in context and to perform the calculation.	15 - 20
	Conjecturing	The ability to formulate appropriate questions and conjectures in order to make sense of quantitative information and to recognise the tentativeness of conjectures based on insufficient evidence.	0 - 5
	Interpreting	The ability to interpret quantitative information (in terms of the context in which it is embedded) and to translate between different representations of the same quantitative information. This interpretation includes synthesising information from more than one source. For example: identifying the correct algebraic formula or graphical representation from a verbal description of a relationship, interpreting the results of a calculation in the original context, deriving and using data from more than one representation in order to solve a problem.	10 - 15
	Reasoning	The ability to identify whether a claim is supported by the available evidence, to formulate conclusions that can be made given specific evidence, or to identify the evidence necessary to support a claim.	5 - 10
	Representing quantitative information	The ability to represent quantitative information verbally, graphically, diagrammatically, and in tabular form using appropriate representational conventions and language. For example: choosing appropriate/correct representations of quantitative information.	5 - 10

More than ten years later we recognise that the specifications might well benefit from revision and refinement after a decade of experience of using them, incorporating new knowledge about the domain gleaned from the literature and our experience in teaching QL. For example, the description of the competence area, “Data representation and analysis” should probably be expanded, given the degree of importance assigned to it.

Table 2
Specification for the Quantitative Literacy Test: Classification Based on Mathematical and Statistical Content

Competence Area	Description/Specifications	% of Test
Quantity, number, and operations	The ability to order quantities, calculate, and estimate the answers to computations required by a contexts using whole numbers, fractions, decimals, percentages, ratios, scientific notation, etc. and simple operations (+, -, ×, ÷, positive exponentiation) on them. The ability to express the same decimal number in alternative ways (such as by converting a fraction to a percentage, a common fraction to a decimal fraction and so on). The ability to interpret the words and phrases used to describe ratios between quantities within a context, to convert such phrases to numerical representations, to perform calculations with them, and to interpret the result in the original context. The ability to work similarly with ratios between quantities represented in tables and charts and in scale diagrams.	25 - 30
Shape, dimension, and space	The ability to understand the conventions for the measurement and description (representation) of 2- and 3-dimensional objects, angles, and direction. The ability to perform simple calculations involving areas, perimeters and volumes of simple shapes such as rectangles and cuboids.	10 - 15
Relationships, pattern, and permutation	The ability to recognize, interpret, and represent relationships and patterns in a variety of ways (graphs, tables, words, and symbols.) The ability to manipulate simple algebraic expressions using simple arithmetic operations.	10 - 15
Change and rates	The ability to distinguish between changes (or differences in magnitudes) expressed in absolute terms and those expressed in relative terms (for example as percentage change). The ability to quantify and reason about changes or differences. The ability to calculate average rates of change and to recognise that the steepness of a graph represents the rate of change of the dependent variable with respect to the independent variable. The ability to interpret curvature of graphs in terms of changes in rate.	10 - 15
Data representation and analysis	The ability to derive and use information from representations of contextualised data in tables (several rows and columns and with data of different types combined), charts (pie, bar, compound bar, stacked bar, “broken” line, scatter plots), graphs, and diagrams (such as tree diagrams) and to interpret the meaning of this information. The ability to represent data in simple tables and charts, such as bar or line charts.	20 - 25
Chance and uncertainty	The ability to appreciate that many phenomena are uncertain and to quantify the chance of uncertain events using empirically derived data. This includes understanding the idea of taking a random sample. The ability to represent a probability as a number between 0 and 1, with 0 representing impossibility and 1 representing certainty.	5 - 10

Table 3
Specification for the Quantitative Literacy Test: Classification Based on Cognitive Processing Levels

Cognitive Processing Levels	Description/Specifications	% of Test
Function 1: Basic Knowledge	<p>Items functioning at the basic-knowledge level require test-takers only to demonstrate knowledge of mathematical and statistical facts, vocabulary and simple algorithms.</p> <p><i>Some examples:</i></p> <ul style="list-style-type: none"> • Calculate using the basic operations including the operations +, -, ×, and ÷ and appropriate rounding of numbers; • Know and use appropriate vocabulary such as equation, formula, bar graph, pie chart, table of values, diameter, radius, mean, median and mode, maximum, probability. • Know and use simple formulae such as the area of a rectangle, where each of the required dimensions is given. • Read information directly from a table, chart or graph. • Know the conventions for representation of numbers and operations such as exponentiation or use of scientific notation. 	15 - 20
Function 2: Applying routine procedures in familiar contexts	<p>Items at the applying-simple-procedures-in-context level require test-takers to perform simple procedures in context. The required procedure is easily identified from the way the problem is posed. All of the information required to solve the problem is immediately available. Little reasoning or interpretation is required.</p> <p><i>Some examples:</i></p> <ul style="list-style-type: none"> • Calculate the price of an article after a given percentage reduction. • Find the ratio between two values read off a chart or table • Use a given scale in a diagram to calculate the dimension of an object represented. • Identify the most appropriate type of graphic representation for a simple set of data. 	25 - 30
Function 3: Applying multi-step procedures in a variety of contexts	<p>Items at the applying-multistep-procedures-in-context level require test-takers to perform procedures that are not immediately obvious and involve more than one step. Some reasoning, interpretation, or synthesis will be required.</p> <p><i>Some examples:</i></p> <ul style="list-style-type: none"> • Select the appropriate data from a chart or several charts and use it to solve a problem or make comparisons. • Identify and perform calculations involving intermediate steps, estimation, or unit conversion. 	25 - 30
Function 4: Reasoning and reflecting	<p>Items at the reasoning-and-reflecting level require test-takers primarily to apply higher-order thinking such as deductive reasoning, synthesizing, and evaluation.</p> <p><i>Some examples:</i></p> <ul style="list-style-type: none"> • Determine the truth or falsity of statements using available evidence. • Evaluate validity of arguments. • Identify the correct graphical representation of a given practical situation involving rates of change. • Generalise patterns observed in situations, make predictions based on these patterns. 	25 - 30

The Structure, Administration, and Scoring of the NBTP QL Test

The NBTP tests are administered at test centres under standardised conditions by specially trained invigilators (proctors). Candidates may take the test twice prior to entering higher education. The items in the QL test consist of two 30-minute sections containing 25 multiple choice items each, making up a total of 50 items. These comprise two out of the seven sections of the NBTP AQL (Academic and Quantitative Literacy) test which is administered in paper and pencil format. The other five sections consist of four academic literacy sections and an unscored pre-test section. Calculators are not used, because it is not possible to assume that all candidates possess them. However, candidates are only required to calculate with simple numbers, for example, with fractions that can easily be simplified by cancellation. Many questions can be answered by estimation.

Each QL test is mapped onto the test specification tables (Tables 1, 2 and 3 above). The tests are available in English and Afrikaans, which are currently the two languages of instruction used in South African higher education institutions.

Items are developed in item-development workshops by panels composed of lecturers, who teach first-year courses in higher education, and a small number of school teachers. These items are reviewed for content representation and fairness and sensitivity to ensure that they do not display possible bias or differential item functioning for the groups of interest. Those items that are approved are selected to be pre-tested as non-scoring items in assembled NBTP AQL tests. After pre-testing, items are reviewed by review panels for their psychometric properties. Items that are found to be acceptable are placed in an item database to be used as regular items. Twenty-five forms of the test are assembled and administered each year and include regular test items, common (anchor) items, and pre-test items (in some forms).

Responses to items are evaluated dichotomously, that is as either right or wrong. These responses are then scored using the unidimensional three-parameter (a, b, c) Item Response Theory (IRT) model, where a = discrimination, b = difficulty, and c = guessing/pseudo-chance (Yen and Fitzpatrick 2006) and a score is generated for each candidate on a scale of 0% to 100%.

Each different form of the QL test contains common items for equating purposes. Modern test theories are used to ensure that the scores on different forms of the tests are linked and equated (Holland and Dorans 2006) to ensure that performance is not a function of the version of the test that the candidate has taken.

The results are reported using criterion-referenced benchmarks set through standard-setting methods to place candidates in one of three proficiency bands. These describe both their preparedness for the demands of higher education and the extent to which the curricula should be responsive to their level of preparedness. Standard-setting workshops to determine the benchmark levels take place every

Table 4
QL Test Performance Standards and Their Interpretations for Bachelors Study (BE) and for Diploma and Higher Certificate Study (DE & CE)

Performance Band	Score Range	Description
Proficient	BE 70%-100%	Test performance suggests that future academic performance will not be adversely affected. (Students may pass or fail at university, but this is highly unlikely to be attributable to strengths or weaknesses in the domains tested.) If admitted, students may be placed into regular programmes of study.
	DE & CE 63%-100%	
Intermediate	BE 38%-69%	The challenges identified are such that it is predicted that academic progress will be adversely affected. If admitted, students' educational needs should be met as deemed appropriate by the institution (e.g., extended or augmented programmes, special skills provision).
	DE & CE 34%-62%	
Basic	BE 0%-37%	Test performance reveals serious learning challenges: it is predicted that students will not cope with degree-level study without extensive and long-term support, perhaps best provided through bridging programmes (i.e., non-credit preparatory courses, special skills provision) or further education and training provision. Institutions admitting students performing at this level would need to provide such support themselves.
	DE & CE 0%-33%	

three years, in which panels of lecturers in first-year courses in higher education take part (last done in October, 2015). The standard-setting method that is employed to determine the benchmarks is the modified Angoff method (Hambleton and Pitoniak 2006). In the modified Angoff method panellists of experts are asked to estimate independently the probability that the borderline candidate (a candidate believed to be located at the borderline of a particular proficiency category) will answer each item correctly. These probabilities are summed over items for each panellist to arrive at a panellist performance standard for that borderline, and then these panellist performance standards are averaged to obtain a performance standard for the panel on the test. The process is repeated for all borderlines between proficiency categories of interest. Basically, the panellists are estimating the expected standard or cut-score or benchmark of the borderline candidate on each item and then these expected item scores are summed to obtain an estimated true score for the borderline candidate on the collection of test items. This process, which involves experts in determining the cut-scores for the proficiency bands, supports the argument for the validity of the use of the test for describing the proficiency of prospective students for the demands of higher education.

Table 4 provides a description of the current QL benchmark levels defining proficiency levels for the three kinds of qualification provided by higher education: first (bachelor) degree (BE), diploma (DE), and higher certificate (CE). It also contains the information that is routinely provided to test-takers to describe their proficiency level and to institutions to explain the appropriate responses required.

Table 5
QL Test Intermediate Performance Standards Split into Upper and Lower and Their Interpretations for Bachelors Study (BE) and for Diploma and Higher Certificate Study (DE & CE)

Intermediate Performance Band	Score Range	Description
Intermediate Upper	BE 54%-69%	Students are likely to need complementary support (additional tutorials, workshops, augmented courses, language intensive work).
	DE & CE 49%-62%	
Intermediate Lower	BE 38%-53%	Students need to be placed in an extended degree programme.
	DE & CE 34%-48%	

Because a large proportion of the scores for prospective students in South Africa are in the Intermediate band, it has been found productive to divide the Intermediate performance band into two, the Intermediate Upper and Lower bands as shown in Table 5. It is important to note that this division was not done through the standard-setting exercise but rather by creating a division halfway between the top and the bottom of the Intermediate band of scores. The recommendations for the educational provision appropriate for a student whose performance is in the Intermediate band are that extended or augmented programmes should be provided. In a context of limited resources faced by many higher education institutions, it is useful to identify which of these students are in most urgent need of support.

Validity and Reliability of the NBT QL Test

Validity refers to the degree to which one can make inferences from the scores obtained on a test. Validity is not an absolute state but rather a collection of evidence indicating that the scores obtained on a test are valid for their intended uses (AERA 2014). Until about 1950, predictive value (correlation) was nearly all that was required to argue for validity (Cronbach 1971). In the early 1950's, a more nuanced and deeper appreciation of validity developed and different types of validity were described. In 1989, Messick argued for validity to be seen as an "integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment ..." (Messick 1989, 13).

In this section we will outline briefly some of the evidence which supports the argument for the validity of the NBT QL test. Firstly, we will address construct and content validity evidence. Then, we provide evidence for predictive validity on the basis of studies done into the relationship between performance on the NBT QL

test and in higher education. Thirdly, we will consider some studies into the relationship between student performance in the National Senior Certificate and on the NBT in order to establish evidence for convergent and discriminant validity.

Construct validity evidence refers to the extent to which a test captures a specific theoretical construct or trait (in this case QL ability), and content validity evidence addresses the alignment between test items and the domain area they are intended to assess (in this case QL for higher education). At the level of the test specifications, evidence for content validity is supported by the fact that the test specifications were designed by the authors who are (and were at the time) experienced in QL teaching, curriculum development, test development, and research. Furthermore, items for the tests are developed annually by South African higher education academics, using the QL test specification as their guide, who also verify that the items' content is aligned to the appropriate aspects of the test specifications. In annual item review workshops, other academics review each item to check a) that the item is aligned with the construct, b) that the item has acceptable psychometric properties² and c) that it does not display 'sensitivity' (when different subgroups of test-takers may experience an item in a manner that would adversely affect their performance).

Predictive validity evidence refers to the degree to which a test accurately predicts a criterion that will occur in the future. The evidence for predictive validity of the NBT QL test is supported by research into the relationship between performance on the NBT QL test and performance in higher education at the course, programme, faculty, and institutional levels. For example, Prince (2016), for a sample of first-time entrants into three different faculties, found a statistically highly significant association between NBT QL performance category and academic standing at the end of 6 years of study. There was a progressive decrease from high to low performance levels (on the NBT QL test) in the proportions of students who had graduated after 6 years of study: 77.19% of students whose QL scores were proficient had graduated, while only 45.97% of those with basic scores had done so. There was a corresponding increase from high to low performance levels in the proportion who had left university without graduating (46.45% of those with basic scores and 5.0% for those with proficient scores had failed). A similar study of first-time entering engineering students (Prince et al. 2017) determined that those with proficient QL scores were twice as likely to graduate in minimum time

² In the context of item difficulty, the p -value is the proportion of test-takers that answered an item correctly and ranges from 0 to 1. (This value should not be confused with a p -value related to hypothesis testing.) A high value (≥ 0.95) means that an item is easy, a low value (≤ 0.25) means that the item is difficult. We seek to have p -values in the range 0.30 to 0.89. The point-biserial correlation (r_{pbis}) is a measure of the discriminating, or differentiating, power of the item and ranges from -1 to 1 . A negative r_{pbis} is indicative of a bad item, as it means that lower-scoring test-takers are more likely than higher scoring test-takers to respond correctly. We seek to have the r_{pbis} values greater than 0.2.

as those who were not proficient (intermediate or basic scores) and that the non-proficient group was three times as likely to fail to graduate at all. At a different university from ours, Rankin et al. (2012) investigated the predictive power of the NBT tests for performance of students in the first class test of a first year economics course. They found that although the NSC Mathematics marks were not correlated with students' economics test performances, the scores for the NBT QL test were significantly correlated with their economics test performance.

Convergent validity evidence refers to the degree to which a test is similar to (converges on or can be replaced by) other cognate tests to which it should theoretically be similar. Discriminant validity evidence, on the other hand, refers to the degree to which a test is not similar to (diverges from) other tests to which it should theoretically not be similar. The Pearson linear correlation coefficient (r) is used to measure the extent of convergent and discriminant validity, a higher value indicating convergence and lower value indicating divergence.

As an example, we will consider only the degree to which the NBT QL test converges with or diverges from the NSC Mathematics examination. The NSC Mathematics subject is cognate with the NBT QL construct, in that mathematical ability and quantitative competence are related. Thus one would expect a degree of convergence. However, there are also very important differences between competence in the discipline of mathematics and in QL, leading one to expect that the NSC Mathematics and NBT QL tests would diverge from each other to some extent.

In a study of the relationship between the NBT QL scores and the NSC Mathematics results of 6,271 students, the correlation coefficient was 0.63, indicating a fair degree of convergent validity (Prince and Frith 2017). In a similar study of the results of 56,662 students, the correlation coefficient was 0.58, which while still indicating some convergence, is closer to the cut-off value of $r = 0.5$ used in Prince (2017).

Nevertheless, the NSC Mathematics result in both studies was shown to be a poor predictor of NBT QL score. There was a wide range of NBT QL scores associated with any particular NSC Mathematics result, with this range being wider the higher the NSC Mathematics result. For example in the Prince and Frith (2017) study, for students who obtained over 80% for Mathematics, the NBT QL scores ranged from less than 25% to nearly 100%. Even a low mark for NSC Mathematics was associated with quite a large range of possible NBT QL scores. For example, for those who obtained NSC Mathematics marks less than 30%, the NBT QL scores ranged from about 15% to 75%. For those students who obtained less than 30% for NSC Mathematics, just over 75% obtained a basic NBT QL score. This proportion decreased fairly linearly as the NSC Mathematics marks increased so that only about 5% of the students with NSC Mathematics above 80% obtained NBT QL scores in the basic band. However, even for those who obtained over 80% for NSC

Mathematics, less than 50% obtained a proficient NBT QL score; the proportion of proficient scores for Mathematics results between 70% and 80% was even smaller at just under 25%. In addition, the mean NBT QL scores were 10 percentage points lower than the mean Mathematics results. These observed differences support the fact that the NBT QL and NSC Mathematics assessments are measuring somewhat different constructs and demonstrates the importance of treating them as distinct and complementary.

Reliability refers to the degree of stability and consistency of test scores. The QL test is examined for reliability as measured by Cronbach's alpha (α), which is frequently used to determine internal consistency. Cronbach's alpha indicates how related the scores on the items are to the construct of interest (in this case, QL). While coefficients with a value of 0.70 or higher have traditionally been considered, reliability values above 0.80 are desirable (Nunally 1978). The reliability of the QL tests is routinely calculated for the various administrations of all test forms and values indicating good reliability are always obtained. For example, the reliability values (α) for QL tests taken in 2017 ranged from 0.87 to 0.89.

How the NBTP QL Test Scores Are Used

The results obtained for the NBTP QL test nationally indicate a severe mismatch between the preparedness of students and higher education provision in terms of QL. For example, in 2016, of the 83,110 prospective applicants to higher education who took the NBTP QL test, only 11.7% performed at the Proficient QL level. According to the benchmarks set through the standard setting process, the remaining 88.3% were expected to experience academic challenges arising from their low levels of QL proficiency if their curricula needs were not met by higher education institutions. Just about 43% performed at the Basic level, which means that they would be severely challenged academically in higher education (Centre for Educational Testing for Access and Placement 2017). Thus, it is clear that institutions could benefit from using the information provided by the NBTP QL test to place students in appropriate programmes and to develop educational interventions to more effectively bridge this "articulation gap" (to use the words of Scott et al. 2007, 42) between the level of many students' QL and the demands of higher education.

For various reasons (briefly discussed later) the NBT QL test data is not used by some of the twenty-six universities in the country and usage is only partial in most of those that use it. Some of the reasons provided for not using the NBTP test scores, or not using them optimally, are the cost to students (R80, about US\$6 in 2017) and the challenges faced by rural students who have to travel to test centres. For some universities, writing the NBTP tests is compulsory and in some institutions it is compulsory for entry into specific faculties (the divisions of

universities comprising related subject areas). The scores are most commonly used for diagnosis to determine appropriate academic support, to inform admission and placement decisions in specified departments and faculties, and to help plan interventions. Informal surveys and interviews have established that the NBTP tests, including the QL test, are used in a variety of ways by the majority of universities. Some institutions use the NBTP test scores as part of their selection criteria, although they were used differently and to different extents by different faculties. About a third of universities use test scores in various ways for placement purposes and some use them for diagnostic purposes.

As an example of how one institution that does use the NBTP data uses the QL scores, we will describe in more detail how they are used at our institution. UCT requires all applicants who are resident in South Africa to take the Academic and Quantitative Literacy NBTP tests, (and for Commerce, Engineering, Health Sciences, and Science they must also take the Mathematics test). The test scores are used in a variety of ways in making admission and placement decisions in the different faculties.

Only Engineering and Health Sciences use the NBTP test scores as part of the admission-points score which is used amongst other criteria for selection purposes, each of the three NBTP test scores (including the QL test) contributing one ninth of the total admission points. In these faculties, one of the criteria to be guaranteed a place is that the applicant achieves Proficient scores in all the NBTP tests and these faculties also do not accept applicants with NBTP test scores in the Basic band. The use of the NBT test scores to inform admissions decisions in these faculties is supported by research carried out internally.

Although other faculties do not use the NBTP test scores in the calculation of admission points, they also use them as a component of their selection criteria. For example, Commerce includes amongst its requirements for admission to its undergraduate degrees that a student would have achieved an NBTP QL test score in the Upper Intermediate band. This requirement is relaxed to Lower Intermediate in some cases for the students admitted from targeted redress groups who are deemed to be from a historically disadvantaged background. In Law, all students must have NBTP QL test scores in the intermediate band or above. On the other hand, the NBTP QL test score is not used to determine entry into Humanities as a whole, though it will determine whether a student can enter certain courses with quantitative requirements, especially Psychology 1. Science is alone in not using any of the NBTP test scores in any way for admissions decisions.

In several faculties, the NBTP test scores are used internally for assisting with decisions about placement in programmes, courses, and teaching interventions. In some cases, the NBTP QL test scores are used formally to determine which courses a student should (or need not) take, and in other cases the scores are used to select students for additional support within courses. For example, in Humanities the

NBTP QL test score is used alongside the school-leaving results to determine if a student can register immediately for first-year Psychology (if their scores determine that they are Proficient in QL), whether they must take QL courses provided by the NC concurrently with Psychology 1 (if their scores determine that they are in the Upper Intermediate proficiency category in QL), or whether they will only qualify to enter Psychology 1 after completing the QL courses (if their scores determine that they are in the Lower Intermediate or Basic proficiency category in QL). On the other hand, Law takes the approach that all its undergraduate students should complete a QL half-course (also provided by the NC) but grants an exemption to those students whose NBTP QL score is in the Proficient category.

Less formally, in the first year of the six-year undergraduate programme for Medical doctors, the NBTP QL test scores are used to determine which students should attend weekly QL workshops (provided by the NC) and which students are deemed capable of mastering the quantitative components of the course independently (namely those whose QL scores are Proficient).

Although Science does not use the NBTP test scores for admission purposes, their students are required to take these tests, and the scores are used as part of the information set informing placement decisions early in students' first year, when they may be advised to transfer to an extended degree programme.

Ideally there would be many other examples across the institution (that we are not aware of) of the NBTP QL test scores being used to help academics and students to make decisions about appropriate curriculum, and so we hope that the examples above are not an exhaustive list.

Discussion and Conclusion

The first goal of the NBT project is to assess the competencies of entry-level students in higher education. During the last nine years, the NBTP QL test has been used to assess the QL of South African school-leavers who take the test in order to apply to institutions of higher education. The results are routinely reported to institutions. In 2017, it is estimated that approximately 40% of 196,391 students entering higher education took the NBTP QL test (Govender 2016). The results have revealed that most of these students are severely underprepared in terms of the QL demands of higher education.

The second goal is to assess the relationship between entry-level proficiencies and school-level exit outcomes, provided by the NSC examination. The NSC is the national school-leaving examination and is not primarily intended to assess the readiness of a school-leaver for higher education. Thus, the NSC and the NBTP assessments are designed for different purposes. In addition, the NSC is norm-referenced while the NBTP tests are criterion-referenced (Prince 2016). As a result, the NSC and NBTP tests provide different but complementary information about a

student's capabilities (Prince 2017). In general, the NSC mathematics subject assessments provide an artificially positive impression of school-leavers' abilities, especially for lower-performing students (Nel and Kistner 2009; Schoer et al. 2010; Rankin et al. 2012). On the other hand the NBTP tests are designed to be realistic about the school-leavers' readiness for the expectations of higher education. For example, comparing the results for a large sample of test-takers revealed that in general the NBTP QL scores were about 9 percentage points lower than the NSC Mathematics scores and the NBT QL scores were about 27 percentage points lower than the NSC Mathematical Literacy results. Less than 50% of those who scored at the highest level for NSC Mathematics and less than 5% of those who scored at the highest level for NSC Mathematical Literacy obtained scores classified as Proficient in the NBTP QL test (Prince and Frith 2017).

Both the results of the assessment of the entry-level QL competency of students and the analysis of the relationship between these NBTP QL scores and students' school-leaving results point to the need for a more comprehensive systemic response, requiring curriculum change at many levels (Prince and Frith 2017). There is, however, a reluctance across the higher education sector to embrace the idea of testing, which is generally viewed with scepticism, and thus to recognise the value of the information provided by the NBTP QL test results. In addition, although educators are generally aware of the importance of academic language literacy for success at university, QL is not as widely recognised as a barrier to learning. For example, the latest UCT strategic framework draft document made no mention of it. The result is that what amounts to a significant QL challenge in higher education is therefore not being addressed at appropriate levels by the higher education sector, nor to our knowledge is it being addressed effectively by individual institutions. At UCT a partial response is provided through the NC's provision of courses and interventions for some students (in Humanities, Law, and Health Sciences).

The third goal of the NBT project is to provide information to institutions that will be useful in making admissions and placement decisions. The NBTP QL data is used in decisions about admissions only by those institutions and programmes that are highly selective. For example, the NBTP test score contributes to the admissions points used for admission to medical doctors' training nationally and to some engineering programmes. These are programmes where a high proportion of the applicants have very strong NSC results, and the additional complementary information provided by the NBTP tests is therefore valued. There is, however, a reluctance generally to use the NBTP test scores as part of the admissions criteria for various political reasons and as a result of conservatism within the system. The NBT project is seen by some to be a competitor with the national education department, which administers the NSC, and an attempt by elite institutions to shoehorn in a university entrance examination. Some students have also criticised

the NBTP tests on the grounds that they are perceived to discriminate against black students (many of whom unfortunately do not perform well on them as a result of the poor educational experience they have been provided with at school). Institutions understandably wish to avoid being seen as elitist or discriminatory, especially in the current climate of student protest and calls for a decolonised higher education system.

These criticisms of the NBTP tests should not affect their usefulness in making decisions about placement of students in appropriate programmes of study, but in fact the NBTP data have not been well-used in informing placement decisions (except in some cases). The use of the test results for placement decisions in many institutions is impractical due to the fact that they are unable to ensure that their students arrive at registration with the tests already taken and the results available. Reasons for student failure to complete the test in advance include the cost to the applicant of writing the test and the fact that some institutions accept late applications. A greater effort may have to be made by the NBT project team to convince university staff who have responsibility for making placement policy decisions of the usefulness of the NBTP test results for this purpose.

The fourth goal of the NBT project, to provide diagnostic information to inform curriculum responses, is the most difficult to realise and a major area requiring development. The NBT project team has held workshops with staff at institutions to explain how the NBTP data might be used to inform curriculum interventions to address QL. These are, however, only partially successful because many disciplinary lecturers are too busy to attend, are not interested, or lack the necessary expertise to translate diagnostic information from testing into meaningful responses in terms of curriculum for developing students' QL. It takes some effort to understand the data and its implications, and developing curriculum responses requires time and an understanding of the nature of QL which most university staff do not have. To provide useful diagnostic data to university teachers and to facilitate their efforts to change their curriculum even if they are willing to do so requires significant input over time from the NBTP team who do not have sufficient resources to do this effectively.

Although the NBT project was commissioned by Higher Education South Africa to assist institutions in addressing urgent and important educational problems, institutions do not display great commitment to the idea of compulsory testing of all applicants. It has also proved difficult for a small team of academics based at one university to overcome the inertia in the system, the general tendency of academics to want to work independently rather than collaborate, and the reluctance of many to recognise that they need to change how they teach to provide students with what they need rather than what the institution traditionally has offered. On the other hand, the project has collected a wealth of very useful data

about the QL of school-leavers and there is thus potential for much of the necessary development work to be enriched by it.

Acknowledgements

We would like to thank the reviewers who made useful suggestions for improvement of this paper.

References

- American Educational Research Association, American Psychological Association & National Council of Measurement in Education. 2014. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Archer, Arlene, Vera Frith, and Robert N. Prince. 2002. "A Project-Based Approach to Numeracy Practices at University Focusing on HIV/AIDS." *Literacy and Numeracy Studies* 11(2): 123–31.
- Baynham, Mike and Dave Baker. "'Practice' in Literacy and Numeracy Research: Multiple Perspectives." *Ways of Knowing* 2(1): 1–9.
- Boaler, J. 2002. "Exploring the Nature of Mathematical Activity: Using Theory, Research." *Studies in Mathematics* 51(1-2): 3–21.
<https://doi.org/10.1023/A:1022468022549>.
- Centre for Educational Testing for Access and Placement. 2017. *National Benchmark Tests Project National Report: 2017 Intake Cycle*. Cape Town, South Africa: Centre for Educational Testing for Access and Placement.
- Chapman, Anne. 1998. "Academic Numeracy: Developing a Framework." *Literacy and Numeracy Studies* 8(1): 99–121.
- Chapman, Anne and Alison Lee. 1990. "Rethinking Literacy and Numeracy." *Australian Journal of Education* 34(3): 277–89.
<https://doi.org/10.1177/000494419003400305>.
- Council on Higher Education. 2017. *VitalStats: Public Higher Education 2015*. Pretoria, South Africa: Council on Higher Education.
- Cronbach, Lee J. 1971. "Test Validation." In *Educational Measurement*, 2nd ed., edited by Robert L. Thorndike, 443–507. Washington DC: American Council on Education.
- Department of Basic Education, South Africa. 2005. *National Curriculum Statement Grades 10–12 (General) Subject Assessment Guidelines: Mathematical Literacy*. Pretoria, South Africa: Department of Education.
- Frith, Vera, Lynn Bowie, Kate Gray, and Robert Prince. 2003. "Mathematical Literacy of Students Entering First Year at a South African University." In *Proceedings of the Ninth National Congress of the Association for*

- Mathematics Education of South Africa, Rondebosch, South Africa*, 30 June–4 July: 186–93.
- Frith, Vera and Geney Gunston. 2011. “Towards Understanding the Quantitative Literacy Demands of a First-Year Medical Curriculum.” *African Journal of Health Professions Education* 3(1): 19–23.
- Frith, Vera and Pam Lloyd. 2016. “Investigating Proportional Reasoning in a University Quantitative Literacy Course.” *Numeracy* 9(1): Article 3. <https://doi.org/10.5038/1936-4660.9.1.3>.
- Frith, Vera and Robert N. Prince. 2006. “Quantitative Literacy.” In *Access and Entry Level Benchmarks: The National Benchmark Tests Project*, edited by Hanlie Griesel, 28–34; 47–54. Pretoria, South Africa: Higher Education South Africa.
- Frith, Vera and Robert N. Prince. 2009. “A Framework for Understanding the Quantitative Literacy Demands of Higher Education.” *South African Journal of Higher Education* 23(1): 83–97. <https://doi.org/10.4314/sajhe.v23i1.44804>.
- Frith, Vera and Robert N. Prince. 2016. “Quantitative Literacy of School Leavers Aspiring to Higher Education in South Africa: Lessons from the South African National Benchmark Quantitative Literacy Test.” *South African Journal of Higher Education* 30(1): 138–61. <https://doi.org/10.20853/30-1-552>.
- Gal, Iddo, Mieke van Groenestijn, Myrna Manly, Mary Jane Schmitt, and Dave Tout. 2005. “Adult Numeracy and Its Assessment in the ALL Survey: A Conceptual Framework and Pilot Results.” In *Measuring Adult Literacy and Life Skills: New Frameworks for Assessment*, edited by T. Scott Murray, Yvon Clermont, and Marilyn Binkley, 137–91. Ottawa, Canada: Statistics Canada.
- Govender, Prega. 2016. “Tensions Simmer Over First-Year University Enrolments for 2017.” *Mail and Guardian*, November 28, 2016.
- Griesel, Hanlie. 2006. *Access and Entry Level Benchmarks: The National Benchmark Tests Project*. Pretoria, South Africa: Higher Education South Africa.
- Hambleton, Ronald K. and Mary J. Pitoniak. 2006. “Setting Performance Standards.” In *Educational Measurement*, 4th ed., edited by Robert L. Brennan, 433–70. Westport, CT: Greenwood/Praeger.
- Holland, Paul W. and Neil J. Dorans. 2006. “Linking and Equating.” In *Educational Measurement*, 4th ed., edited by Robert L. Brennan, 187–220. Westport, CT: Greenwood/Praeger.
- Hughes-Hallett, Deborah. 2001. “Achieving Numeracy: The Challenge of Implementation.” In *Mathematics and Democracy: The Case for*

- Quantitative Literacy*, edited by Lynn Arthur Steen, 93–98. Washington, DC: The National Council on Education and the Disciplines.
- Jablonka, Eva. 2003. “Mathematical Literacy.” In *Second International Handbook of Mathematics Education*, edited by Alan J. Bishop, M. A. Clements, Christine Keitel, Jeremy Kilpatrick, and Frederick K. S. Leung, 75–102. Dordrecht, The Netherlands: Kluwer Academic Publishers.
https://doi.org/10.1007/978-94-010-0273-8_4.
- Kelly, Sheilagh, Betty Johnston, and Mike Baynham. 2007. “The Concept of Numeracy as Social Practice.” In *The Adult Numeracy Handbook: Reframing Adult Numeracy in Australia*, edited by Sheilagh Kelly, Betty Johnston, and Keiko Yasukawa, 35–49. Sydney, Australia: Adult Literacy and Numeracy Australian Research Consortium.
- Messick, Samuel. 1989. “Validity.” In *Educational Measurement*, 3rd ed., edited by Robert L. Linn, 13–103. Englewood Cliffs NJ: Prentice Hall.
- Mullis Ina V. S., Michael O. Martin, Teresa A. Smith, Robert A. Garden, Kelvin D. Gregory, Eugenio J. Gonzalez, Stephen J. Chrostowski, and Kathleen M. O’Connor. 2003. *Trends in International Mathematics and Science Study (TIMSS): Assessment Frameworks and Specifications 2003*, 2nd ed. Boston, MA: International Association for the Evaluation of Educational Achievement and International Study Center, Lynch School of Education.
- National Assessment Governing Board. 2004. *Mathematics Framework for the 2005 National Assessment of Educational Progress*. Washington, DC: US Department of Education.
- Nel, C. and L. Kistner. 2009. “The National Senior Certificate: Implications for Access to Higher Education.” *South African Journal of Higher Education* 23(5): 953–73. <https://doi.org/10.4314/sajhe.v23i5.48810>.
- Nunnally, Jum C. 1978. *Psychometric Theory*, 2nd ed. New York: McGraw-Hill.
- OECD. 2003. *The PISA 2003 Assessment Framework—Mathematics, Reading, Science, and Problem-Solving Knowledge and Skills*. Paris, France: OECD.
- Prince, Robert N. 2016. “Predicting Success in Higher Education: The Value of Criterion and Norm-Referenced Assessments.” *Practitioner Research in Higher Education* 10(1): 22–38.
- Prince, Robert N. 2017. “The Relationship between School-Leaving Examinations and University Entrance Assessments: The Case of the South African System.” *Journal of Education* 70: 133–60.
- Prince, Robert N. and Arlene Archer. 2006. “Quantitative Literacy as Situated Social Practice in Higher Education.” *International Journal of Learning* 12(7): 227–33. <https://doi.org/10.18848/1447-9494/CGP/v12i07/47938>.
- Prince, Robert N. and Arlene Archer. 2008. “A New Literacies Approach to Academic Numeracy Practices in Higher Education in South Africa.”

- Literacy and Numeracy Studies* 16(1): 63–75.
<https://doi.org/10.5130/lms.v16i1.1948>.
- Prince, Robert N. and Vera Frith. 2017. “The Quantitative Literacy of South African School-Leavers Who Qualify for Higher Education.” *Pythagoras* 38(1): a355. <https://doi.org/10.4102/pythagoras.v38i1.355>.
- Prince, Robert N., Vera Frith, and Jacob J. Jaftha. 2004. “Mathematical Literacy of Students in First Year of Medical School at a South African University.” In *Proceedings of the 12th Annual Conference of the Southern African Association for Research in Mathematics, Science and Technology Education*, edited by Andy Buffler and Rudiger C. Laugksch, 791–98. Durban, South Africa: SAARMSTE.
- Prince, Robert N., Darlington Mutakwa, and Janine Dunlop. 2017. “School-Leaving and University Entrance Assessments in Explaining Performance in Engineering Studies.” In *Proceedings of the Fourth Biennial Conference of the South African Society for Engineering Education, Cape Town, South Africa* 14- 15 June 2017.
- Prince, Robert N., and Zach Simpson. 2016. “Quantitative Literacy Practices in Civil Engineering Study: Designs for Teaching and Learning.” In *Proceedings of the 5th International Conference on Designs for Learning*, edited by Anne-Mette Nortvig, Birgitte Holm Sørensen, Morten Misfeldt, Rikke Ørngreen, Benjamin Brink Allsopp, Birgitte Henningsen, and Heidi Hautopp, 189–204. Aalborg Universitetsforlag.
- Rankin, Neil, Volker Schoer, C. Sebastiao, and C. van Walbeek. 2012. “Predictors of Academic Performance: National Senior Certificate Versus National Benchmark Test.” *South African Journal of Higher Education* 26(3): 564–85.
- Schoer, Volker, Miracle Ntuli, Neil Rankin, C. Sebastiou, and K. Hunt. 2010. “A Blurred Signal? The Usefulness of National Senior Certificate (NSC) Mathematics Marks as Predictors of Academic Performance at University Level.” *Perspectives in Education* 28(2): 9–18.
- Scott, Ian, Nan Yeld, and Jane Hendry. 2007. *Higher Education Monitor No. 6: A Case for Improving Teaching and Learning in South African Higher Education*. Pretoria, South Africa: The Council on Higher Education.
- Steen, Lynn Arthur. 2001. “The Case for Quantitative Literacy.” In *Mathematics and Democracy, The Case for Quantitative Literacy*, ed. L. A. Steen, 93–98. Washington, DC: The National Council on Education and the Disciplines.
- Steen, Lynn Arthur. 2004. *Achieving Quantitative Literacy: An Urgent Challenge for Higher Education*. Washington D.C.: The Mathematical Association of America.

- Street, Brian. 2005. "Applying New Literacy Studies to Numeracy as Social Practice." In *Urban literacy: Communication, Identity and Learning in Development Contexts*, edited by Alan Rogers, 87–96. Hamburg, Germany: UNESCO Institute for Education.
- Street, Brian and Dave Baker. 2006. "So, What about Multimodal Numeracies?" In *Travel Notes from the New Literacy Studies*, edited by Kate Pahl, and Jennifer Rowsell, 219–33. Clevedon, England: Multilingual Matters Ltd. <https://doi.org/10.21832/9781853598630-012>.
- Yen, Wendy. M. and Anne R. Fitzpatrick. 2006. "Item Response Theory." In *Educational Measurement*, 4th ed., edited by Robert L. Brennan, 111–53. Westport, CT: Greenwood/Praeger.