




# Using standard setting to promote meaningful use of mathematics assessment data within initial teacher education programmes



## Authors:

Qetelo M. Moloi<sup>1</sup>   
Anil Kanjee<sup>1</sup>   
Nicky Roberts<sup>2</sup> 

## Affiliations:

<sup>1</sup>Department of Primary Education, Tshwane University of Technology, Pretoria, South Africa

<sup>2</sup>Centre for Education Practice Research, University of Johannesburg, Johannesburg, South Africa

## Corresponding author:

Qetelo Moloi,  
qetmoloi@gmail.com

## Dates:

Received: 27 Apr. 2019  
Accepted: 09 Nov. 2019  
Published: 12 Dec. 2019

## How to cite this article:

Moloi, Q.M., Kanjee, A., & Roberts, N. (2019). Using standard setting to promote meaningful use of mathematics assessment data within initial teacher education programmes. *Pythagoras*, 40(1), a493. <https://doi.org/10.4102/pythagoras.v40i1.493>

## Copyright:

© 2019. The Authors. Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

## Read online:



Scan this QR code with your smart phone or mobile device to read online.

Within initial teacher education there is increasing pressure to enhance the use of assessment data to support students to improve their knowledge and skills, and to determine what standards they meet upon graduation. For such data to be useful, both programme designers and students require meaningful and comprehensive assessment reports on students' performance. However, current reporting formats, based on percentages, are inadequate for providing meaningful qualitative information on students' mathematics proficiency and, therefore, are unlikely to be used for interventions to improve teaching and enhance learning. This article proposes standard setting as an approach to reporting the assessment results in formats that are meaningful for decision-making and efficacious in subsequent interventions. Mathematics tests, developed through the Primary Teacher Education (PrimTEd) project, were administered electronically on a convenient sample of first-year and fourth-year PrimTEd students ( $N = 1\,377$ ). The data were analysed using traditional descriptive statistical analysis and the Objective Standard Setting method. The two reporting formats – one using a percentage score and the other using standards-based performance levels – were then compared. The study identified important distinguishing features of students' mathematical proficiency from the two reporting formats, and makes important findings on the specific knowledge and skills that students in South African initial teacher education programmes demonstrate. We conclude that reporting assessment results in standards-based formats facilitates differentiated interventions to meet students' learning needs. Furthermore, this approach holds good prospects for benchmarking performance across universities and for monitoring national standards.

**Keywords:** Standard setting; meaningful reporting; Objective Standard Setting; performance levels; performance standards.

## Introduction

The centrality of assessment to effective and deeper learning in higher education has been asserted in a number of research studies (Coates & Seifert, 2011; Filetti, Wright, & William, 2010; Ping, Schellings, & Beijaard, 2018; Russell & Markle, 2017; Sambell, McDowell, & Montgomery, 2013). While there has been some research conducted on the value, and use, of feedback from assessment by students at university level (Sadler, 1998), no corresponding research has been reported on meaningful formats of reporting results of tests and examinations in ways that can help lecturers improve teaching and help enhance learning by students. Arguing that assessment remains one of the final frontiers of change, Coates (2015) notes that despite the substantial improvement within many parts of the higher education sector, assessment systems have not changed for a century, and that student knowledge and skills are still most commonly measured in traditional ways. However, there has been recognition of the limitations of reporting assessment results in raw scores, including percentages, and the tendency to aggregate scores of exclusive and independent constructs (Coates & Seifert, 2011). These limitations present a challenge to the users of assessment results because the extent to which lecturers and students can interact with assessment results, in ways that can help them learn better, hinges heavily on the meaningfulness, the consistency and comprehensiveness of the assessment reports that they receive.

To address this challenge, a new project was established aimed at changing the dominant culture from one in which assessment data is only utilised for reporting, promotion and certification to

one in which assessment data is regarded as a rich source of information for use in improving teaching and learning. Meaningful reporting and utility of assessment data is a central requirement for programme coordinators of Bachelor of Education programmes who are collaborating in the assessment workstream of the Primary Teacher Education (PrimTEd) project. The aim of the PrimTEd assessment workstream is to:

- Advocate for Higher Education Institutions' participation in common assessment approaches.
- Encourage collaboration on teacher assessment approaches towards developing teacher competence in relation to improving the teaching and learning of mathematics.
- Develop a teacher competency assessment framework (related to mathematics), for Foundation and Intermediate Phase, for student teachers graduating from Initial Teacher Education (ITE) programmes.

This is not a trivial task as there is, as yet, no agreement about the knowledge needed for teaching mathematics. Even in countries such as the United States, where there are claims of teaching standards that have been developed, and teachers are assessed against such standards using a variety of assessments, review of released assessment items reveals that there is still 'lack of agreement over what teachers need to know' (Hill, Shilling, & Ball, 2004, p. 12).

In the South African context such agreement is being sought through the collaborative work of three PrimTEd workstreams (that focus on mathematical thinking, number sense and geometry and measurement). Teams of professionals from across South African universities are collaborating on assessing the mathematics knowledge of student teachers when they enter the Bachelor of Education (B.Ed.) programme in their first year, and again when they exit at fourth-year level. The purpose of agreeing on such a common assessment – and reporting on its findings – is to allow programme designers and lecturers to reflect on and improve their programmes over time.

The article reports on observations made in the PrimTEd assessment workstream (with a specific focus on mathematics). Firstly, the shortcomings of current reporting practices are presented. The article explores and illustrates the value of the standard setting approach for improving the reporting and use of assessment data. Next, the article presents the process followed in identifying performance levels (PLs), performance level descriptors (PLDs), and cut-scores to determine students' levels of proficiency (Kanjee & Moloji, 2016). This is followed by a presentation of available results to demonstrate how a standards-based framework can be used to establish and report the levels at which students are functioning in order to: (1) identify specific learning needs of students for developing appropriate interventions to address these needs and (2) determine the knowledge and skills with which students graduate upon completion of their ITE programmes. The article concludes by outlining implications for enhancing the use of assessment data to improve ITE programmes for

mathematics teachers in South Africa and proposing the next steps in developing a teacher competency assessment framework for universities participating in the PrimTEd.

## Situating this study

There has been an international shift in focus in teacher education towards an 'intensification of standardisation', which in many countries has resulted in high stakes summative assessment. Summative assessment is then expected to fulfil multiple functions: providing evidence of a teacher candidate's achievement at a given point against the standards, guiding and informing jurisdictional decision-making around teacher preparation, licensure and advanced certification, and furnishing evidence of teacher effectiveness and readiness to teach (Allen, 2017; DeLuca & Klinger, 2010).

Allen (2017) argues that focusing on summative assessment is revealing of what is valued and privileged – and what is not – in teacher education. In contrast – although not as dichotomies – formative assessment in teacher education begins with students understanding what it is they are intending to learn, offering feedback to students, student goal setting and keeping track of their own learning and includes formative use of summative tests (Brookhart, 2017). As such there is a growing requirement within the field of ITE to design standardised tests which can be meaningful to both programme designers and to students. Yet in a systematic review of what, how, and why teacher educators learn (Ping et al., 2018), a focus on assessment and how to use it to improve teacher education practice appears to be missing.

It is therefore unsurprising that in South Africa, reporting on teacher knowledge for mathematics has been limited and – as evident in the studies discussed below – reporting on assessment has been in relation to mean attainment of percentage of teachers meeting a benchmark. In one empirical study that involved practising teachers, Venkat and Spaull (2015) analysed the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) data collected from a representative sample of learners and their teachers ( $n = 401$ ) to determine levels of teachers' content knowledge. The authors used the Curriculum and Assessment Policy Statement grade level and content strands to classify the SACMEQ test items to better align to the South African curriculum, and 60% as a minimum benchmark for mastery at the Grade 6/7 level. Venkat and Spaull reported that 79% of South African Grade 6 mathematics teachers were classified as having content knowledge levels below Grade 6.

There was even less research available about student teachers in ITE programmes. A scan of literature relating to 'initial teacher education' and 'assessment' in South Africa revealed a dearth of studies with this focus. A recent systemic review of South African studies pertaining to ITE for primary school, and particularly Foundation Phase (Baxen & Botha, 2016), identified these main thematic areas: diversity and social

justice, theory and practice, curriculum concerns, identity and subjectivity, language issues and policy and transformation. The type, utility and reporting on assessment is conspicuous by its absence, and is not mentioned even as a subsection to any of the themes. Baxen and Botha (2016, p. 10) describe the ITE research landscape in South Africa as 'nascent' with the research projects reviewed considered 'isolated and uncoordinated'.

Taylor (2018) refers to a 2010 Council on Higher Education report which described the state of the ITE sector as uncoordinated and of unknown quality of performance. Deacon (2012) called for the establishment of benchmarks – to diagnose what mathematics and English the students entering the B.Ed. programmes possess. The Initial Teacher Education Research Project (ITERP) considered intermediate phase courses in five universities and found that the quality of ITE in South Africa, in terms of the knowledge and skills possessed by first-year entrants, the content that is offered and the exit competencies of final-year students, was questionable in relation to mathematics and language courses. Bowie and Reed (2016, p. 116) reported that all five lecturers in the ITERP study noted that on entering a B.Ed. programme some student teachers are no more proficient in mathematics and English than the Grade 4 to 6 learners they are preparing to teach. In their review of PrimTED assessment results for first-year students ( $n = 1\ 117$ ) across seven universities, Alex and Roberts (2019, p. 68) reported that the 'majority of the first-year students (71%) do not meet the minimum benchmark of 60% for knowledge of the mathematics content at primary school level'.

All of these studies – pertaining to both practising teachers and those in ITE programmes – simply report a deficit in mathematical knowledge for teaching. The mean results and percentage of those attaining a minimum 'benchmark' provide quantitative evidence of a dire situation but do not offer qualitative information on what may be done or how best to intervene.

### Typical, and additional, forms of reporting assessment results

Traditionally assessment results are reported largely in raw scores which are mainly presented as percentage correct responses obtained by students in a test or examination. In any typical report emanating from end-of-year examinations or large-scale assessments, results are generally presented as mean percentage scores by subject areas, often accompanied by some reporting scale that provides an interpretation of the scores. For example, 60% and greater is considered as mastery level (Alex & Roberts 2019; Venkat & Spaul 2015). In some instances (for example, as in Fonseca, Maseko, & Roberts, 2018), additional information is also reported on the performance distribution of students taking the tests, as well as by sub-domain and cognitive levels for the subject tested (see Figure 2 and Figure 5 discussed in a later section). This information may be useful for monitoring student progress or identifying students that meet required levels of performance.

However the utility of this information for identifying specific learning needs of students, or determining the specific sets of knowledge and skills that students master, is limited (Bond & Fox, 2007; Montgomery & Connolly, 1987). Raw scores on their own cannot be used qualitatively to indicate what students were able or not able to do in a test (Bond & Fox, 2007; Moloi & Kanjee, 2018). Raw scores, including percentages, are sample- and test-dependent. On the one hand, in the same test, a sample of relatively more proficient students will achieve higher percentage scores than their less-proficient counterparts. On the other hand, the same sample of students will achieve lower scores in a more difficult test than on a relatively easier one.

Furthermore, the use of percentage scores is predicated on the assumption that the effort of improving one's score on the lower end of the proficiency continuum is the same on the upper end. Yet, studies based on the latent test theory show that it is more difficult to improve from a score of, say, 90% to 95% than from 20% to 25% even though the interval (5%) is the same in both cases (Bond & Fox, 2007). These weaknesses render the use of percentage scores in reporting performance generally unreliable as a measure of either student proficiency or progression over time. As a result, performance standards which are seen to be more stable and provide detailed information on what students know and can do are increasingly used (Glass & Hopkins, 1984).

### Our focus

This article explores and proposes standard setting as an approach to reporting the results of assessment in formats that are meaningful for decision-making and efficacious in subsequent interventions. The premise of the article is that, unless assessment data is analysed and reported in meaningful formats, it is unlikely to be used for interventions that will lead to improvement in teaching and to enhancing learning by students.

Importantly, this article does not seek to examine what teachers of primary school mathematics are required to know, and hence the underlying constructs and possible assessment standards. Such research is being conducted across multiple mathematics workstreams within PrimTED. Rather this article seeks to demonstrate how existing assessment data may be used to feed into the broader research agenda, and offer reporting templates that are both meaningful and usable for participating universities.

### Value and use of standards-based reporting of assessment results

Setting performance standards involves determining a cut-score that separates test-takers into identifiable categories on the basis of what they know and can do, or what they do not know and cannot do, in a test. Along the continuum from not knowing or not being able to do anything at all in a given test, on the one end, to the other extreme of knowing or being able to do everything, what score is necessary for performance

to be judged as 'acceptable'? To determine the necessary score, one has to establish 'a set of content standards and a set of test questions intended to reflect that content as a starting point to a process that will lead directly to setting performance standards' (Barton 1999, p. 19).

Cizek and Bunch (2007, p. 13) define standard setting as 'a process of establishing one or more cut-scores on a test for purposes of categorising test-takers according to the degree to which they demonstrate the expected knowledge and/or skills that are being tested'. The standard setting process involves technical analysis of student responses as well as expert inputs from teams of professionals and members from relevant stakeholder groups who serve to validate the technical results (Tiratira, 2009).

Performance standards specify the amount of knowledge or skills that a test-taker must demonstrate, on a given continuum separated by cut-scores, to satisfy the demands of a particular outcome. An important feature of performance standards is PLs and PLDs. Zieky and Perie (2006, p. 3) describe PLs as general policy statements that indicate the official position on the desirable number and labels of categories to be used in classifying students according to their knowledge and skills in a particular subject and grade. PLDs are defined as detailed descriptions of 'the knowledge, skills, and abilities to be demonstrated by students who have achieved a particular performance level within a particular subject area' (Zieky & Perie, 2006, p. 4). Morgan and Perie (2005, p. 5) affirm that PLDs are 'working definitions of each of the performance levels (that) ... define the rigor associated with the performance levels.'

At the university level, PLDs can provide detailed information on the specific knowledge and skills with which students enter and leave higher education programmes. At individual student level, lecturers can identify students who meet, or do not meet, expected levels of knowledge or skill, categorise them according to need, and decide on appropriate interventions for each category. Not only can performance standards help focus interventions to address specific individual needs, but they can also ensure that students and lecturers share a common understanding of what is expected in student performance (California Department of Education, 2007, p. 2). Clear knowledge of what is expected may encourage students, lecturers and universities to exert more effort to achieve it. Performance standards also help improve the precision with which reporting to relevant stakeholders can be made. Clearly and precisely communicated student progress reports which are based on fair, valid, and easy-to-understand standards demystify expectations and increase trust between the concerned institution and its clients (California Department of Education, 2007, p. 2).

While different methods of developing performance standards and corresponding PLDs have been explored, a common conclusion from research has been that each method has its own strengths and weaknesses and, therefore, the best standard setting method may not exist (Tiratira, 2009).

Notwithstanding possible pitfalls related to the complexities and subjectivity inherent in different methods of setting performance standards, Cizek and Bunch (2007, p. 36) and Bejar (2008, p. 3) emphasise the critical role of communicating and using clear standards to enhance decision-making among all key stakeholders.

## The purpose of the study

The purpose of this study was to explore the use of the standard setting approach for reporting of PrimTED mathematics results so as to enhance their use by lecturers to identify and address specific learning needs of students. In this regard, the following research questions guided the study:

- What reporting formats provide meaningful information to identify the support first-year B.Ed. students require for improving their mathematics knowledge and skills?
- What reporting formats provide meaningful and valid information on the mathematical knowledge and skills that final-year students graduate with upon completion of their ITE programme?

## Research methodology

The study involved testing student teachers' knowledge and skills in mathematics, followed by a standard setting exercise to select the number of PLs, establish the PLDs and calculate the cut-scores that categorised students according to identified levels of performance (Cizek & Bunch, 2007), and the analysis and reporting of results. Adopting a cross-sectional design, the same PrimTED assessment was administered at both first-year and fourth-year levels in this study.

## Sample

A purposive sample for this study was drawn from eight universities that participated in the administration of the PrimTED mathematics test to either their first-year or fourth-year students or both in 2018. A total of 1377 students from the eight universities participated in the testing, of which 85% were students in first year and 15% were in fourth year (Table 1). The students' mathematics test data was subsequently obtained from the PrimTED coordinators. In order to maintain confidentiality and to protect the identities of the participating universities, no demographic data is given.

**TABLE 1:** Student sample participating in the 2018 PrimTED assessment.

University	First year (2018)		Fourth year (2018)		Total	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
A	251	21.4	0	0.0	251	18.2
B	72	6.1	0	0.0	72	5.2
C	70	6.0	87	43.1	157	11.4
D	124	10.6	0	0.0	124	9.0
E	232	19.7	30	14.9	262	19.0
F	238	20.3	0	0.0	238	17.3
G	0	0.0	85	42.1	85	6.2
H	188	16.0	0	0.0	188	13.7
<b>Total</b>	<b>1175</b>	<b>100.0</b>	<b>202</b>	<b>100.0</b>	<b>1377</b>	<b>100.0</b>

## Test design and administration

As it is not the direct focus of this article, we only briefly describe the design and administration of the PrimTED mathematics test. Fonseca et al. (2018) describe the test design pilot, refinement process and how it was administered online across universities. Further details as well as exemplar items are presented in Alex and Roberts (2019). For this article we simply highlight some key features of the test, understanding of which will help to interpret the two report formats – percentage score reporting and standards-based reporting – which we contrast later in the article.

We now know that the ‘content knowledge for teaching mathematics consists of more than the knowledge of mathematics held by any well-educated adult’ (Hill et al., 2004, p. 27). The overarching construct that the PrimTED test began to assess was ‘mathematical knowledge for teaching: knowledge teachers require to teach primary mathematics well’. Alex and Roberts (2019) argue that mathematical knowledge for teaching includes knowledge of: (1) the content (know *how* to do the mathematics itself), (2) why the content makes sense, (3) how to represent concepts using multiple *representations*, (4) how the particular aspect of content *connects* to other topics and grades, and (5) at what *stage* children are ready to learn this content.

The PrimTED mathematics test was a first attempt at designing a common assessment framework for use across ITE programmes. The development of the test was done in the absence of agreed assessment standards (which are being developed across the PrimTED mathematics workstreams). As such, we do not claim that the PrimTED mathematics test assesses all aspects of mathematical knowledge for teaching in primary schools, but the above description is used to elaborate on the test intention and its long-term objective. In this article we simply refer to ‘mathematical knowledge and skills’ to describe the underlying construct that the current PrimTED test begins to assess.

Fonseca et al. (2018) explain that the PrimTED mathematics items were classified as lower cognitive demand, higher cognitive demand or pedagogy items. The cognitive demand levels applied the Stein, Grover and Henningsen (1996) framework on tasks, where ‘lower cognitive demand’ items were considered to be routine procedures; the ‘higher cognitive demand’ items involved moves between representations, required insight, connected across topic areas or had no obvious procedure or starting point (Venkat, Bowie, & Alex, 2017). An example of the distinction between lower and higher cognitive demand is offered in Figure 1.

The third category of items related to ‘mathematical pedagogy’ and were specifically phrased to solicit analysis of a student’s work or of a common error in mathematics (Alex & Roberts, 2019, p. 65). This category of knowledge is described as ‘pedagogical content knowledge’ by Shulman (1986, 1987), and is thought to ‘include familiarity with topics children find interesting or difficult, the representations most

useful for teaching an idea, and learners’ typical errors and misconceptions’ (Hill et al., 2004, p. 12–13). An example item to assess pedagogy is presented in Figure 2.

The PrimTED mathematics test was administered online to first-year and fourth-year students and was monitored by at least one mathematics lecturer at each university. The instrument comprised 50 multiple choice mathematics items which covered four of the mathematics topics prescribed in the Curriculum and Assessment Policy Statement document for intermediate phase (DBE, 2011). These include numbers, operations and relationships, patterns, functions and algebra, space and shape (geometry) and measurement. Data handling was excluded as this topic forms a very small component of the foundation and intermediate phase curriculum at primary school level. The answers to all items were either ‘multiple choice’ or ‘single answer’. No partial marking was necessary. Each item was marked correct (1) or incorrect (0) using online marking.

## Ethical considerations

This study was undertaken following the ethical processes requiring voluntary, informed consent for educational research and under the University of Johannesburg’s protocol number of 2017-072. This ethics application was for the

Rational number, low cognitive demand	Rational number, high cognitive demand
0.7 is a decimal fraction. Write 0.7 as a common fraction.	A farmer’s cost for milk production is R3.12 for each litre. What is his production cost for 2.5 litres of milk? The calculation you need, to get the correct answer is: A. $3.12 \times 2.5$ B. $3.12 - 2.5$ C. $2.5 \div 3.12$ D. $3.12 \div 2.5$

Source: Fonseca, K., Maseko, J., & Roberts, N. (2018). Students mathematical knowledge in a Bachelor of Education (Foundation or Intermediate Phase) programme. In R. Govender & K. Junqueira (Eds.), *Proceedings of the 24th Annual National Congress of the Association for Mathematics Education of South Africa* (pp. 124–139). Bloemfontein: AMESA (p. 129)

FIGURE 1: Exemplar items for lower and higher cognitive demand.

**Rational number, Pedagogy**

Linda says that  $2\frac{3}{4} = \frac{5}{6}$  and she uses the figure below to show why she says so.

Choose the BEST explanation of the key fault in Linda’s reasoning

A. She doesn’t know how to represent fractions

B. She is using different units to represent the whole part and the proper fractions in  $2\frac{3}{4}$

C. She doesn’t know what a mixed number is.

D. She doesn’t know that  $2\frac{3}{4}$  is not  $\frac{5}{6}$  but is actually  $\frac{11}{4}$

Source: Alex, J., & Roberts, N. (2019). The need for relevant initial teacher education for primary mathematics: Evidence from the Primary Teacher Education Project in South Africa. In N. Govender, R. Mudaly, T. Mthethwa, & A. Singh-Pillay (Eds.), *Proceedings of the 27th Conference of the Southern African Association for Research in Mathematics, Science and Technology Education* (pp. 59–72). Durban: SAARMSTE (p. 65)

FIGURE 2: Exemplar pedagogy item.

PrimTED assessment process as a whole and included all participating universities. First-year and fourth-year B.Ed. students were invited by their mathematics course coordinators to write the PrimTED mathematics test, as a voluntary part of their B.Ed. programme. They were told that the assessment was not 'for marks' but rather to be used as diagnostic assessment. On commencement of the online test, each student could choose to 'opt in' to have their data from the test included in a wider research study and for publication, or to 'opt out' and have their data excluded. They were assured that all data would be anonymised.

## Development of performance levels and performance level descriptors

A standard setting session was convened at which mathematics lecturers from the participating universities defined the four PLs (see Figure 3), established the PLDs and participated in the process of determining the cut-scores. The Objective Standard Setting (OSS) method (Stone, 1995) was used to determine cut-scores. The choice of the OSS method was informed by our research that indicates this method as being both cost-effective and providing curriculum-rich information for meaningful reporting of assessment results. Interested readers can refer to Moloï and Kanjee (2018) for general procedures of standard setting and to Stone (2001) for specific technical procedures of the OSS method. The information gathered from the standard setting session was used to conduct an intensive audit of the knowledge and skills that characterise students at each

performance level to establish the PLDs. This was intended to be illustrative of process and did not involve all of the participating university lecturers. It is acknowledged that further discussion relating to these PLDs is required for common agreement on these to be reached. A summary of the illustrative PrimTED mathematics test PLDs is given in Figure 4.

For relative ease and accuracy of reporting we proposed the use of four reporting levels defined as *Not Achieved*, *Partly Achieved*, *Achieved* and *Advanced*. The process of the development of PLs, their policy definitions as well as their generic descriptions has been presented in detail in Moloï and Kanjee (2018). The definition of PLs includes information on the implications of the PL in terms of student progression and necessary intervention (Figure 3).

Figure 4 provides a summary of the PrimTED PLDs with illustrative knowledge and skills that characterise a student's mathematics proficiency at each PL.

The distinctions made in the opening clause of each level descriptor draw on the components of mathematical proficiency as advocated by Kilpatrick, Swafford and Findell (2001, p. 5):

- Conceptual understanding: comprehension of mathematical concepts, operations, and relations.
- Procedural fluency: skill in carrying out procedures flexibly, accurately, efficiently, and appropriately.

Level	Level description	Progression implications	Intervention implications
Advanced	Performance at this level indicates that a student demonstrates <b>comprehensive understanding</b> of the knowledge and skills required to function at this level	Student has high likelihood of success in the next year	Student requires little or no academic intervention but needs to be provided with more challenging tasks to maximise their full potential
Achieved	Performance at this level indicates that a student demonstrates <b>sufficient understanding</b> of the knowledge and skills required to function at this level	Student has a reasonable likelihood of success in the next year	Student may require some assistance with complex concepts to progress to the advance level
Partly achieved	Performance at this level indicates that a student demonstrates <b>partial understanding</b> of the knowledge and skills required to function at this level	Student unlikely to succeed in the next year without support	Student requires specific intervention to address knowledge gaps, while also requiring additional support to progress to the required Achieved level
Not achieved	Performance at this level indicates that a student demonstrates <b>very limited</b> understanding of the knowledge and skills required to function at this level	Student unlikely to succeed in the next year without significant support	Student requires specific intervention to address knowledge gaps, while also requiring additional teaching time and extensive and continued support to progress to the required Achieved level

Source: Moloï, M., & Kanjee, A. (2018). Beyond test scores: A framework for reporting mathematics assessment results to enhance teaching and learning. *Pythagoras*, 39(1), a393 (p. 13). <https://doi.org/10.4102/pythagoras.v39i1.393>

FIGURE 3: Generic performance level definitions.

Partly achieved	Achieved	Advanced
<p>Student mostly shows evidence of 'procedural fluency' as opposed to 'conceptual understanding' and can:</p> <ul style="list-style-type: none"> <li>• do simple straightforward operations +, -, × and ÷ that involve whole numbers.</li> <li>• identify whole numbers on a number line.</li> <li>• identify regular geometric shapes.</li> <li>• calculate area and perimeter of rectangles using numbers, not symbols.</li> <li>• solve problems of one variable, e.g. time, money.</li> </ul>	<p>Student shows evidence of using 'procedural fluency' and some evidence of conceptual understanding and can:</p> <ul style="list-style-type: none"> <li>• operate equally well with symbols and numbers.</li> <li>• make reasonable estimations of spatial dimensions and have 'good sense' of proportion.</li> <li>• express decimals as common fractions and vice versa, and do estimations that involve both.</li> <li>• solve complex problems that involve more than one variable, e.g. money and mass.</li> <li>• support their viewpoints with valid reasons.</li> </ul>	<p>Student functions at 'procedural fluency', 'conceptual understanding', 'adaptive reasoning' and 'strategic competence' and can:</p> <ul style="list-style-type: none"> <li>• use operations to support reasoning.</li> <li>• organise and arrange numbers, variables and functions in logical order to solve problems.</li> <li>• work efficiently with a wide spectrum of real numbers.</li> <li>• 'visualise' and operate complex spatial transformations to solve problems.</li> <li>• support their viewpoints with valid reasons.</li> <li>• analyse learners' work and show awareness of common learner errors.</li> </ul>

FIGURE 4: Illustrative PrimTED mathematics performance level descriptors.

- Strategic competence: ability to formulate, represent, and solve mathematical problems.
- Adaptive reasoning: capacity for logical thought, reflection, explanation, and justification.

While these are used as intertwined strands of mathematical proficiency, the analysis of items likely to be correct at each performance level, together with the analysis of expert contributions to the OSS process, was suggestive of these distinctions. These PLDs are only intended to be illustrative. As noted above, a detailed process of engagement with all mathematics workstreams on this process is still required to ensure common understanding and utility of these kinds of descriptors.

In Figure 4 the PLDs for the *Not Achieved* level, that is, the level below the *Partly Achieved* level, have been omitted. Typically, students who function at this level are characterised by scores that fall below the lowest cut-score and tend to provide sporadic responses that follow no discernible pattern of proficiency.

The hierarchy of PLs means that the knowledge and skills at a given PL are cognitively more demanding than those at a lower PL and less demanding than those at a higher PL: *Not Achieved* < *Partly Achieved* < *Achieved* < *Advanced* in terms of cognitive demand. Secondly, a student who displays knowledge and skills that are characteristic of a given PL is expected to also have a high probability of displaying knowledge and skills at lower PLs but is unlikely to display knowledge and skills at higher PLs. For instance, a student who functions at the *Achieved* level is expected to also demonstrate the knowledge and skills at *Partly Achieved* level but is unlikely to display *Advanced* level knowledge and skills.

## Analysis

Analysis was conducted at four levels. Firstly, the PrimTED mathematics test was investigated for appropriateness of 'targeting', which is an indication of whether the spread of item difficulties appropriately matched the abilities or proficiencies of the students (Wright & Stone, 1979). Tests that are not properly targeted have large measurement errors and low reliability indexes (Bond & Fox, 2007, p. 43) and this may affect standards that are set from such tests. Secondly, the ratings of items from the standard setting process were analysed, using the OSS method (Stone, 2001) to determine cut-scores that separate different PLs. Thirdly, using the cut-scores, students were categorised based on their test scores into one of the four PLs. This categorisation indicated the proportions of first-year and fourth-year students who functioned at each PL. Lastly, statistical analyses were conducted, using descriptive tests, t-tests and chi-square tests to determine significant differences where comparisons of scores were made. The results are presented as examples of both mean score reporting and performance-based reports to illustrate the benefits and effects of using each reporting format to improve teaching and enhance learning.

## Findings and discussion

The findings of the study are presented and discussed in this section. Firstly, the item map was generated to determine the validity of the PrimTED test as an instrument that will produce consistent information about the test-takers' proficiency when other relevant factors remain unchanged (Newby, 2014). Next the cut-scores derived from using the OSS method are presented. Thirdly, practical implications of applying the standards-based approach to address each research question are presented. In order to illustrate the value of this approach results of mean score reporting are also provided for each question addressed. Finally, a summary of the comparisons and their implications is presented.

### PrimTED item targeting

The item map that provides a visual representation of how well the PrimTED test item difficulties or cognitive demands matched the proficiencies or abilities of the test-takers is presented in Figure 5.

As reflected in Figure 5, the easiest mathematics item was Item 8 with Item 22, Item 1 and Item 26 relatively more difficult but overall on the easy end of the cognitive spectrum (although there were still a few students who had a 50-50 probability of answering these items correctly). It is worth noting that there were up to 24 test-takers, represented by the two dots at the bottom end of the map, for whom all the items were too difficult, that is, they had less than 50% chance of answering any of the items correctly. Item 50, Item 38 and Item 47 were on the difficult end of the spectrum but again there were still a few students who had a 50-50 probability of answering these items correctly. There were up to 60 test-takers, represented by the six dots at the top end of the map, for whom there was more than 50% probability of answering all the items correctly. The rest of the test items, from Item 23 on the easy end to Item 45 on the difficult end, were well matched to the abilities of the test-takers. On the whole, the item map shows that there was a reasonable match between person abilities and item difficulties, suggesting that the PrimTED math test was well targeted to the population of the test-takers. According to Bond and Fox (2007), optimal information about the proficiency of test-takers is obtained when the test is neither too easy nor too difficult for them.

### Cut-scores and their implications

Using the OSS method, the cut-scores that mark transition from one performance level to the next were calculated to be: 40% for the *Partly Achieved* level, 53% for the *Achieved* level and 68% for the *Advanced* level. These cut-scores were used to delineate the PLs and devise the reporting format to provide information on the student's level of functioning as well as the knowledge and skills that the student had mastered. In the next section, the key research questions are addressed.

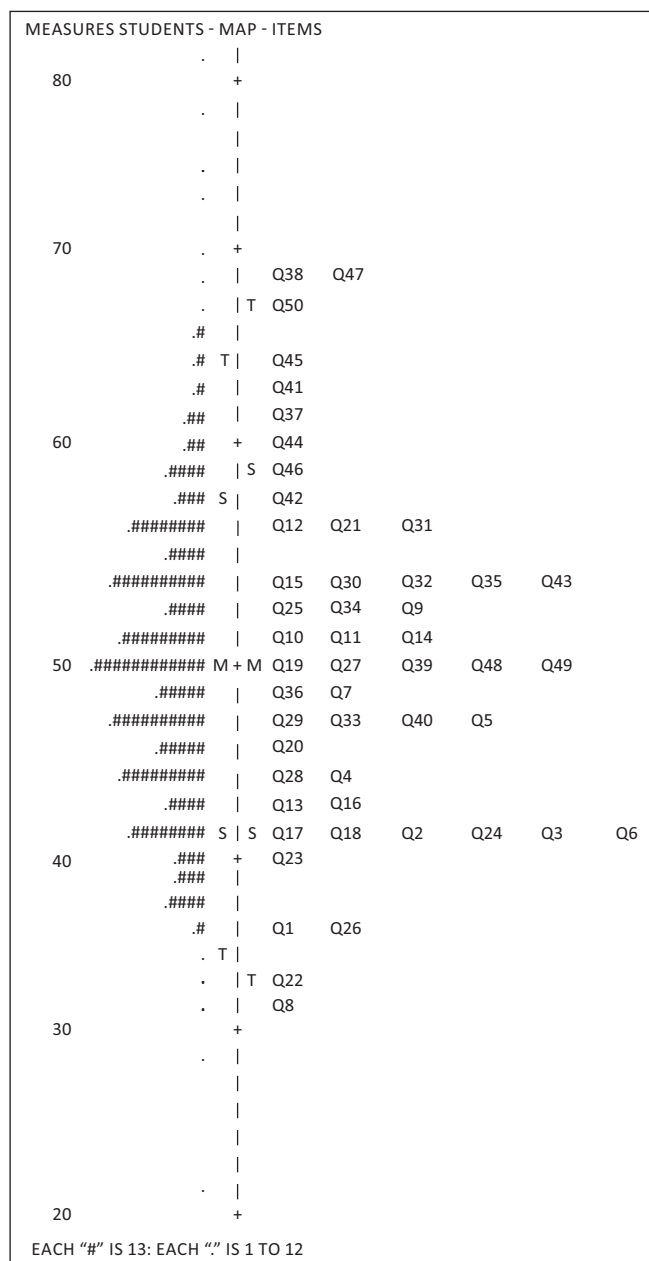


FIGURE 5: PrimTEd mathematics item map.

### Question 1: What reporting formats provide meaningful information to identify the support first-year B.Ed. students require in terms of improving their mathematics knowledge and skills?

Figure 6a and Figure 6b present a percentage-based PrimTEd mathematics report for first-year students from two universities, University A and University B. The information presented in the report comprises: (1) sample size, mean scores and standard deviations, (2) students' score distributions by deciles, (3) student mean scores by content domain and (4) student mean scores by cognitive demand and pedagogy categories.

The results of the independent sample t-test reveal that the mean scores of students in University B are significantly

higher than those of students in University A ( $t(310) = 6.19$ ,  $p < 0.00$ ). Moreover, the mean scores of the content sub-domain areas of the two universities were also consistently higher for University B than for University A. The scores in the two universities differ in three other important aspects: variability, skewness and content cognitive demand. Firstly, University B scores are more variable or spread out ( $SD = 18.24$ ) than those of University A ( $SD = 12.67$ ) signalling greater diversity in mathematics proficiency. Secondly, University B scores are skewed more to the right or towards higher scores signalling the presence of high performing outliers among the test-takers. Thirdly, students in University B were more proficient than their counterparts in University A on items that placed greater cognitive demands on the test-takers.

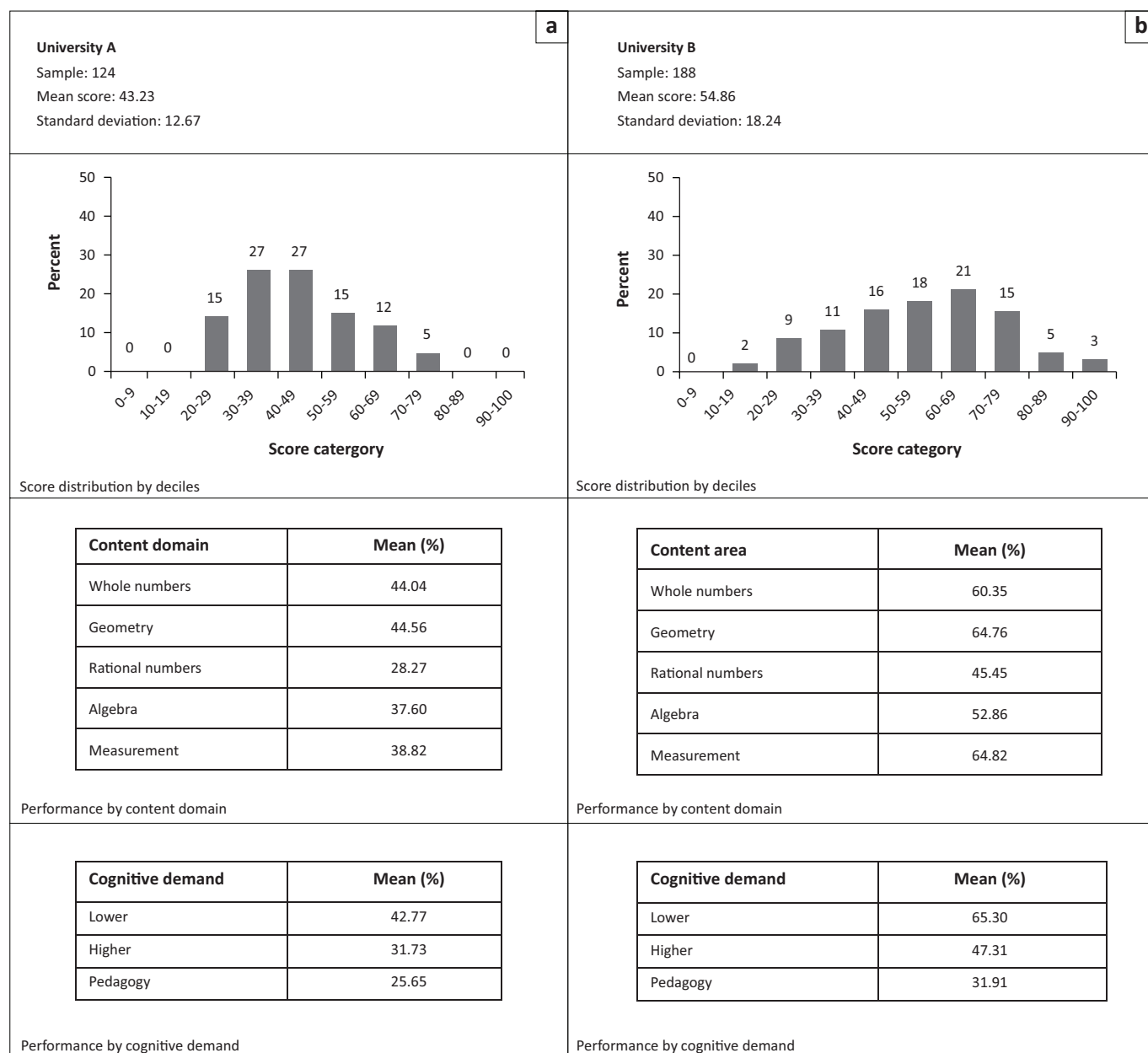
While the aforesaid quantitative descriptive features of relative performance across the universities underscore the distinctiveness of the two universities, they do not, however, provide qualitative information that can be utilised to address specific learning needs of either individual students or categories of students with distinct learning needs. Typically, reports of this kind constrain end-users to employ one-size-fits-all interventions with all known shortcomings of such approaches (Van Geel et al., 2019).

Reflection on the relative strengths relating to particular topics (sub-domains of mathematics arranged by topic in the Curriculum and Assessment Policy Statement) are also evident. From Figure 6a and Figure 6b the content knowledge, measured in mean score percentages, of students in University B were consistently higher than those of their counterparts in University A. This was the case in the overall mathematics scores as well as in the various sub-domains that constitute the subject. Notable was the observation that in both universities the highest mean percentage (whole number) scores were in the sub-domains of Measurement, Whole Numbers and Geometry, ranging between 39% for Measurement and 45% for Geometry in University A, and between 60% for Whole Numbers and 65% for Measurement in University B. For both universities, the fact that Geometry and Measurement were scored relatively higher than other sub-domains may require further investigation as lecturers participating in the team reported that students tend to have more challenges in these sub-domains, particularly in Geometry.

In addition, across both universities the lowest performance was in Rational Numbers and Algebra. Mean scores in these two sub-domains ranged between 28% and 38% in University A and between 46% and 53% in University B. This is an observation of great concern because rational numbers and algebra do not only constitute the majority of the content in secondary school mathematics, but algebraic expressions and rational numbers find a lot of application in the other sub-domains.

Typically, for both universities, student mean scores were higher for items that made lower cognitive demands than





**FIGURE 6:** (a) First-year student 2018 performance for University A; (b) First-year student 2018 performance for University B.

items of a higher cognitive demand. Figure 6a and Figure 6b show that the respective mean scores (rounded to whole numbers) were 43% and 32% for University A compared to 65% and 47% for University B. For both universities the lowest mean scores were in items classified as 'pedagogy', namely 26% in University A and 32% in University B. However, given that the students were in their first year of a four-year course, the low scores in pedagogy may not be of particular concern but they do point to knowledge deficits that need to be addressed.

To adequately equip the student teachers, the percentage score report shows that both universities seem to face a singular challenge of establishing a deeper understanding in the fundamental mathematics content knowledge, although the challenge is much more acute in University A. However, the raw score report does not provide necessary detail on

the specific challenges that each student faced – whether it be in a specific content sub-domains, different cognitive demands or pedagogic methods. For effective and targeted interventions, both lecturers and students would be better equipped if detailed information is provided on what students know, and can or cannot do, in the subject of concern.

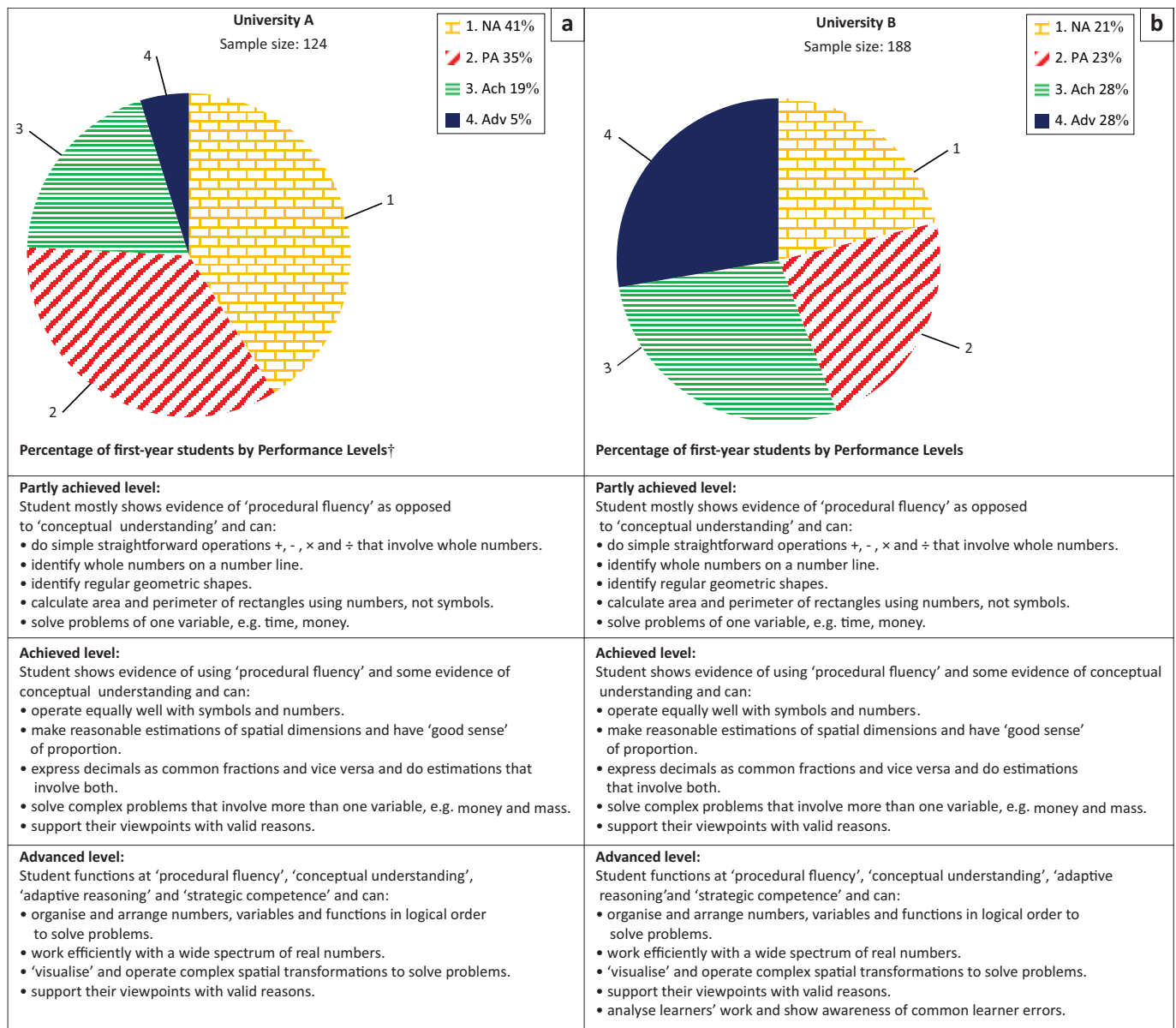
As can be observed from Figure 6a and Figure 6b on the one hand, and Figure 7a and Figure 7b on the other hand, there are markedly distinct features that distinguish the two reporting formats. Figure 7a and Figure 7b present user-friendly information on how students were distributed across the PLs in terms of their mathematics proficiency, as well as information on the knowledge and skills that students who function at a particular PL demonstrate. Also, the reports provide an overall picture of student performance in the form of pie charts. Test results reported in this format are

more likely to be used by lecturers to intervene to improve the performance of their students. Also, this report provides information that the student can use to make informed decisions on how they can improve their performance in specific areas of knowledge and skills.

Figure 7a indicates that about a quarter (24%) of first-year students in University A were functioning at either the *Achieved* or *Advanced* level in mathematics which, according to the appropriate PLDs, are levels characterised by a predominance of conceptual over procedural reasoning. In University B it was just more than half (56%) of the first-year cohort who were functioning at either the *Achieved* or *Advanced* levels. In terms of the definitions of the PLs, these were students who could be expected to succeed and progress to the next level in mathematics with minimal support. In contrast, 41% of students in University A were functioning at

the *Not Achieved* level compared to 21% in University B. This category of students demonstrates very limited knowledge of the requisite mathematics knowledge and skills and is unlikely to succeed without intensive support. Overall, these results indicate that lecturers in University A would have to focus more on supporting students with basic mathematics knowledge and skills than their counterparts in University B.

The typical standards-based report that we recommend in Figure 7a and Figure 7b comprises a number of uses and benefits to both lecturers and students. Firstly, the lecturer who receives this cohort of students, for example mid-year or at the beginning of the second year of study, is provided with a detailed performance profile of the cohort of students and information on what each student knows and can do. Working from the report, the lecturer will be able to design and plan targeted interventions that address the specific



Note: †, For the sake of clarity, we include the illustrative performance level descriptions, as outlined in Figure 4. This provides an indication of the reporting detail which lecturers and students would receive following the PrimTEd test. To avoid unnecessary repetition, in example reports presented later in this article, the PLD details are omitted.

FIGURE 7: (a) First-year student 2018 performance for University A, (b) First-year student 2018 performance for University B.

learning needs of the various categories of students. Secondly, provided that information is available, lecturers are easily able to monitor individual students' progress over time. Thirdly, lecturers can address issues of equity within their classes and ensure that additional and adequate resources and support are provided to those students that have the greatest need, that is, students at the *Not Achieved* and *Partly Achieved* levels.

Fourthly, because each student will know the performance level at which they function, they would be better able to plan and pace their own learning from an informed position in terms of the specific areas of knowledge and skill that they need to focus on. Studies that have contrasted the use of marks or test scores with textual comments in feedback to students have confirmed that students tend to focus on test scores rather than paying attention to comments that direct them to what to do in order to improve their performance (Nicol & Macfarlane-Dick, 2006). Consequently, giving feedback in the form of test scores tends to militate against students becoming self-regulated lifelong learners when compared to using descriptive feedback that outlines how students should improve their own learning (Nicol &

Macfarlane-Dick, 2006). The effect of these benefits of standard setting formats of reporting should be improved and effective teaching that is likely to enhance students' learning.

## Question 2: What reporting formats provide meaningful and valid information on the mathematical knowledge and skills that final-year students graduate with upon completion of their ITE programme?

The mathematics results of fourth- and final-year students' performance for University C and University D have been summarised in Figure 8a and Figure 8b. The percent-score results show that at the point of graduation, the mathematics mean score of students from University C ( $\bar{x} = 51.95\%$  (15.12)) is significantly higher ( $t(179) = 3.27, p < 0.00$ ) than that of their counterparts in University D ( $\bar{x} = 43.79\%$  (17.39)) and further interrogation of the distribution of scores shows that University D has more outlier students of outstanding mathematical proficiency than University C. The outlier students of outstanding proficiency from the apparently weaker University D, in terms of mean scores, account for the

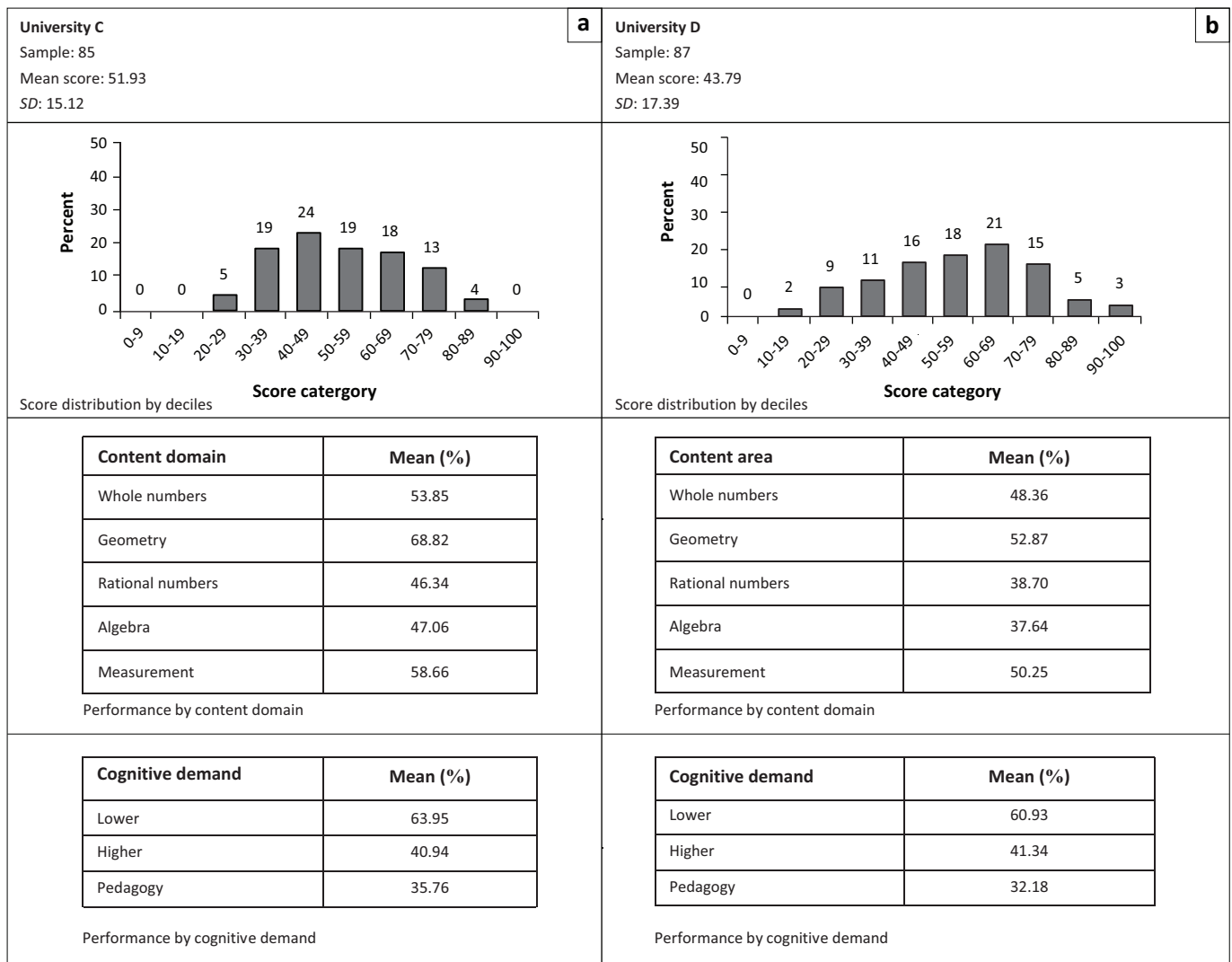


FIGURE 8: (a) Fourth-year student 2018 performance for University C, (b) Fourth-year student 2018 performance for University D.

equal mean scores ( $\bar{x} = 41\%$ ) on items that make high cognitive demands on students in both universities.

In both universities the highest mean percentage scores of students were in the sub-domains of Geometry, Measurement and Whole Numbers, ranging between 54% for Whole Numbers and 69% for Geometry in University C and between 48% for Whole Numbers and 53% for Geometry in University D. A general observation regarding the results for both universities is that students graduated and commenced teaching with far lower knowledge and skills, in terms of their mean scores in the PrimTED test. These scores differ substantially from the minimum of 60% that Venkat and Spaul (2015) applied for determining mastery levels in the SACMEQ data, and which was also adopted by Alex and Roberts (2019). This was the case for both content knowledge and the pedagogical content knowledge where mean scores were 36% for University C and 32% for University D.

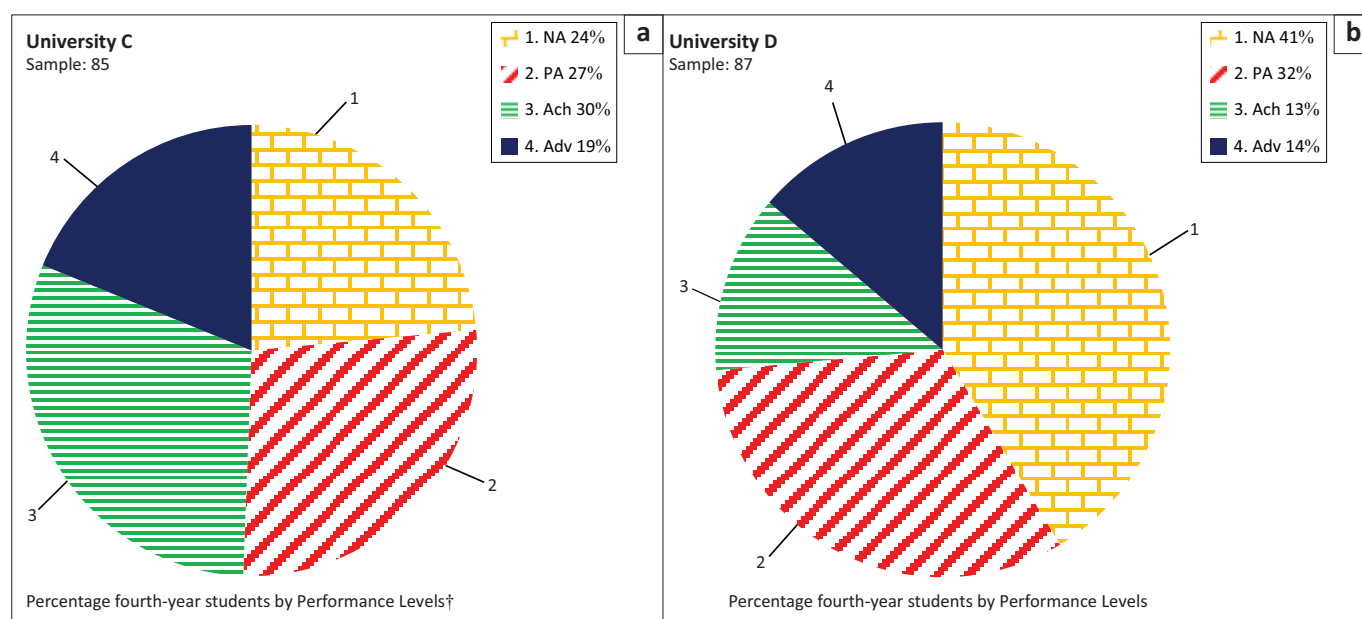
Although detailed analysis of the mean scores distribution serves to unravel inequalities in performance across the universities, it still does not provide information on what the test-takers know, can or cannot do in the subject. The standards-based report, for the same institutions, as shown in Figure 9a and Figure 9b addresses this shortcoming. A review of Figure 9a and Figure 9b indicates that 51% of students from University C and 73% of students from University D may graduate while still functioning at the *Not Achieved* and *Partly Achieved* levels in mathematics.

In terms of the PL definitions and PLDs, these students demonstrate either very limited or partial understanding of the mathematics knowledge and skills required to teach mathematics, and show more evidence of procedural fluency

than conceptual understanding. For instance, they can do the four basic operations arithmetically, but they experience difficulty when they have to work with algebraic symbols. Approximately half of the fourth-year students in University C, compared to 27% in University D, function at either the *Achieved* or *Advanced* level in mathematics. In terms of the PL definitions, they demonstrate either sufficient or comprehensive knowledge or skills to teach mathematics. In terms of PLDs, they show evidence of 'procedural fluency', 'conceptual understanding', 'adaptive reasoning' and 'strategic competence'. For instance, they operate equally well with numbers and symbols, can solve complex problems that involve more than one variable in mathematics and can support their viewpoints with valid reasons. This detail on the students' proficiency and thus their readiness to teach mathematics is very explicit in the PLDs of the standards-based reporting format.

More importantly, the standards-based report provides a clear indication of: (1) how many students can be considered to have graduated from a specific ITE programme with the requisite knowledge and skills, and thus are likely to effectively teach mathematics upon entering the teaching profession, (2) whether the specific ITE programme is fulfilling its mandate to prepare students with the requisite knowledge and skills to teach in their subject areas, and (3) the specific areas that need improvement, within an ITE programme, in order to better prepare students for entering the teaching profession. This information is missing from the mean score reports, and may result in misleading interpretations on how well ITE programmes are functioning.

The key findings from this study were that the traditional use of mean scores to report assessment results lacked necessary information for identifying what students know and can do



Note: †, The detailed illustrative performance level descriptors are presented in Figure 4, and in Figure 7a and Figure 7b. While these are not repeated here, they would be included in reports sent to universities and to individual students.

FIGURE 9: (a) Fourth-year student 2018 results by Performance Levels for University C, (b) Fourth-year student 2018 results by Performance Levels for University D.

and this compromises the feasibility of meaningful and targeted interventions for improving teaching and enhancing learning. Furthermore, this study has shown that the use of PLs to report student performance provides meaningful reporting that facilitates its use by students, lecturers and university management to enhance student learning as well as monitor student progress over time and the impact of the ITE programmes on producing teachers with the requisite content knowledge and skills to effectively teach their subject areas. In this regard, the study addresses a key challenge confronting the higher education sector that Coates and Seifert (2011) identified – that research contributions over the last decade have mainly focused on providing criteria on how to evaluate the impact of such reporting rather than on what needs to be done differently to overcome the weaknesses of aggregating and reporting assessment results in raw scores.

## Conclusion and implications

The ITERP study reported on by Bowie (2014) was an important step towards drawing attention to the lack of common frameworks and assessment of mathematics content knowledge of new primary and secondary school teachers in South Africa. Its findings demonstrated the need for the PrimTEd intervention. The subsequent articles emerging from the PrimTEd assessment workgroup provided a common assessment framework which confirmed the widespread deficits across a wider range of ITE programmes. The overall results of this study cohere with previous reports (Alex & Roberts, 2019; Bowie 2014; Fonseca et al., 2018) in that they make explicit the current deficits in mathematical knowledge for teaching which are evident at both first-year and fourth-year levels in ITE programmes for primary teachers.

More importantly, this study – which draws on the same PrimTEd assessment data – not only highlights the limitations inherent in current reporting approaches, but also proposes an alternative for enhancing the reporting and use of student's assessment results. Specifically, this study has extended the reporting on assessment results beyond mean percentage scores to information-rich, qualitative descriptions of the typical knowledge and skills that characterise student teachers of mathematics in ITE programmes. By enhancing the quality of reporting through the use of performance level reporting, the current study does not only help to fathom the depth of the deficit in mathematics knowledge among student teachers, but it also contributes a response to Deacon's (2012) call for the establishment of benchmarks to diagnose knowledge gaps in student teachers of mathematics. Proactively, institutions will be able to use performance level reports to develop relevant interventions to address specific learning needs of students in mathematics classes, set improvement targets for themselves, and also assist individual students to do the same. The use of PLs in reporting will also provide evidence-based benchmarking across institutions which in turn will create a basis for professional collaboration among staff from the different institutions.

The PrimTEd project provides both the data that can be used among universities that have diverse histories, and the opportunity for exploring ways of monitoring the effectiveness of teacher training. In particular, our study opens vistas for effective ways of reporting assessment results using standards-based formats to overcome the weaknesses inherent in traditional formats of reporting results. To further the PrimTEd research agenda, the PrimTEd assessment workstream will initiate a process to seek agreement on the PLDs. We envisage a process of detailed engagement with the professional standards emerging from across the mathematics workstreams. This includes consideration for the assessment framework and test items. Refining the standards, the assessment instruments used to measure these standards and the related PLDs will require collaboration – working across universities and across the PrimTEd workstreams.

We think such a collaborative process will contribute to securing meaningful buy-in from university lecturers on the preferential use of standards-based reporting. While this article offers an illustration of how this may be done, the university lecturers themselves will be engaged further on the performance level descriptions and the report format. We consider such buy-in and meaningful engagement to be essential for the effective use of the information that flows from such reports to help students learn better.

The implications of adopting a common standards-based reporting format does not only hold good prospects for professional collaboration among universities and the lecturers, but it also creates opportunity for establishing benchmarks for monitoring the effectiveness of teacher development in South Africa. With common standards-based benchmarks it will be possible to provide inter-institution monitoring and support so that there will be common expectations on the knowledge and skills that teacher trainees are armed with when they start their teaching careers. This area forms the basis for further research within the PrimTEd project. Finally, there could be value in providing standards-based reports to schools and thus allowing schools and mentors of newly graduated teachers to use the reports for supporting the students after they enter the teaching profession.

## Acknowledgements

### Competing interests

We declare that we have no financial or personal relationships that might have inappropriately influenced our writing of this article.

### Authors' contributions

This article was developed collaboratively by the three authors. Q.M.M. did the initial literature review, led in technical aspects related to standard setting and prepared the manuscript to comply with publication standards. A.K. led the conceptualisation of the article, prepared the data for analysis, analysed the data and guided the writing of the manuscript. N.R. described the PrimTEd test design process

and prior reporting on the test outcomes, framed the test theoretically in relation to how the instrument was developed, reviewed the level descriptors and led in the quality assurance of the manuscript.

## Funding information

This publication was made possible with the support of: (1) the Teaching and Learning Development Capacity Improvement Programme (TLDCIP) which is being implemented through a partnership between the Department of Higher Education and Training and the European Union, and (2) the Assessment for Learning Niche Area, located within the Faculty of Humanities at the Tshwane University of Technology.

## Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any affiliated agency of the authors.

## References

- Alex, J., & Roberts, N. (2019). The need for relevant initial teacher education for primary mathematics: Evidence from the Primary Teacher Education Project in South Africa. In N. Govender, R. Mudaly, T. Mthethwa, & A. Singh-Pillay (Eds.), *Proceedings of the 27th Conference of the Southern African Association for Research in Mathematics, Science and Technology Education* (pp. 59–72). Durban: SAARMSTE. Retrieved from [https://www.saarmste.org/images/docs/Uploads\\_190224/SAARMSTE%202019%20-%20Long%20Paper%20Proceedings%20FINAL%2031%20Dec%20for%20printing%20PDF.pdf](https://www.saarmste.org/images/docs/Uploads_190224/SAARMSTE%202019%20-%20Long%20Paper%20Proceedings%20FINAL%2031%20Dec%20for%20printing%20PDF.pdf)
- Allen, J. (2017). Summative assessment in teacher education. In D. Clandinin & J. Husu (Eds.), *The SAGE handbook of research on teacher education* (pp. 910–926). London: Sage.
- Barton, P.E. (1999). *Too much testing of the wrong kind: Too little of the right kind in K-12 education*. Educational Testing Service. Retrieved from <https://files.eric.ed.gov/fulltext/ED430052.pdf>
- Baxen, J., & Botha, L.J. (2016). Establishing a research agenda for Foundation Phase initial teacher education: A systematic review. *South African Journal of Education*, 36(3), 1–15. <https://doi.org/10.15700/saje.v36n3a1263>
- Bejar, I.L. (2008). *Standard setting: what is it? Why is it important?* Princeton, NJ: Education Testing Services. Retrieved from [https://www.ets.org/Media/Research/pdf/RD\\_Connections7.pdf](https://www.ets.org/Media/Research/pdf/RD_Connections7.pdf)
- Bond, T.G., & Fox, C.M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). London: Lawrence Erlbaum.
- Bowie, L. (2014). *The Initial Teacher Education Research Project: Report on mathematics courses for intermediate phase student teachers at five universities*. Johannesburg: JET Education Services. Retrieved from [https://www.jet.org.za/resources/copy\\_of\\_bowie-report-on-maths-courses-offered-at-5-case-study-institutions-18-feb.pdf](https://www.jet.org.za/resources/copy_of_bowie-report-on-maths-courses-offered-at-5-case-study-institutions-18-feb.pdf)
- Bowie, L. & Reed, Y. (2016) How much of what? An analysis of the espoused and enacted mathematics and English curricula for intermediate phase student teachers at five South African universities. *Perspectives in Education*, 34(1), 102–119.
- Brookhart, S. (2017). *How to give effective feedback to your students* (2nd ed.). Alexandria, VA: ASCD.
- California Department of Education. (2007). *Development of performance level descriptors for the California standards tests (CSTs) and high school exit exam (CAASSEE)*. Retrieved from <https://www.cde.ca.gov/ta/tg/ca/caapld.asp>
- Cizek, G.J., & Bunch, M.B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. London: Sage.
- Coates, H. (2015). Assessment of learning outcomes. In A. Curaj, L. Matei, R. Precopie, J. Salmi, & P. Scott (Eds.), *The European Higher Education Area – 2015: Between critical reflections and future policies* (pp. 399–413). London: Springer. [https://doi.org/10.1007/978-3-319-20877-0\\_15](https://doi.org/10.1007/978-3-319-20877-0_15)
- Coates, H., & Seifert, T. (2011). Linking assessment for learning, improvement and accountability. *Quality in Higher Education*, 17(2) 179–94. <https://doi.org.10.1080/13538322.2011.554308>
- Deacon, R. (2012). *The Initial Teacher Education Research Project: The initial professional development of teachers: A literature review*. Johannesburg: JET Education Services. Retrieved from <https://www.jet.org.za/resources/deacon-initial-professional-development-of-teachers-literature-review-feb15web.pdf>
- DeLuca, C., & Klinger, D.A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419–438. <https://doi.org/10.1080/0969594X.2010.516643>
- Department of Basic Education. (2011). *Curriculum and assessment policy statement*. Pretoria: DBE. Retrieved from [https://www.education.gov.za/Curriculum/CurriculumAssessmentPolicyStatements\(CAPS\).aspx](https://www.education.gov.za/Curriculum/CurriculumAssessmentPolicyStatements(CAPS).aspx)
- Filetti, J., Wright, M., & William, M.K. (2010). Grades and ranking: When tenure affects assessment. *Practical Assessment, Research & Evaluation*, 15(14), 1–6. Retrieved from <http://pareonline.net/getvn.asp?v=15&n=14>
- Fonseca, K., Maseko, J., & Roberts, N. (2018). Students mathematical knowledge in a Bachelor of Education (Foundation or Intermediate Phase) programme. In R. Govender & K. Junqueira (Eds.), *Proceedings of the 24th Annual National Congress of the Association for Mathematics Education of South Africa* (pp. 124–139). Bloemfontein: AMESA. Retrieved from <http://www.amesa.org.za/AMESA2018/Volume1.pdf>
- Glass, G., & Hopkins, K. 1984. *Statistical methods in education and psychology* (2nd ed.). Boston, MA: Allyn and Bacon.
- Hill, H., Shilling, S., & Ball, D. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, 105(1), 11–30.
- Kanjee, A., & Moloi, Q. (2016). A standards-based approach for reporting assessment results in South Africa. *Perspectives in Education*, 34(4), 29–51. <https://doi.org/10.18820/2519593X/pie.v34i4.3>
- Kilpatrick, J., Swafford, J., & Findell, F. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington DC: National Research Council.
- Moloi, M., & Kanjee, A. (2018). Beyond test scores: A framework for reporting mathematics assessment results to enhance teaching and learning. *Pythagoras*, 39(1), a393. <https://doi.org/10.4102/pythagoras.v39i1.393>
- Montgomery, P.C., & Connolly, B.H. (1987). *Norm-referenced and criterion-referenced tests: Use in paediatrics and application to task analysis of motor skill*. Retrieved from <http://ptjournal.apta.org/content/67/12/1873.long>
- Morgan, D.L., & Perie, M.P. (2005). *Setting cut-scores for college placement*. New York, NY: College Board Research Report Number 2005-9. Retrieved from <https://files.eric.ed.gov/fulltext/ED562865.pdf>
- Newby, P. (2014). *Research methods for education* (2nd ed.). London: Routledge. <https://doi.org/10.4324/9781315758763>
- Nicol, D.J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and several principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Ping, C., Schellings, G., & Beijaard, D. (2018). Teacher educators' professional learning: A literature review. *Teaching and Teacher Education*, 75, 93–104. <https://doi.org/10.1016/j.tate.2018.06.003>
- Russell, J., & Markle, R. (2017). *Continuing a culture of evidence: Assessment for improvement. Research report*. Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/ets2.12136>
- Sadler, D.R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Policy and Practice*, 5(1), 77–84. <https://doi.org/10.1080/0969595980050104>
- Sambell, K., McDowell, L., & Montgomery, C. (2013). *Assessment for learning in higher education*. New York, NY: Routledge. Retrieved from <http://insight.cumbria.ac.uk/id/eprint/3907/>
- Shulman, L.S. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. Retrieved from <http://www.jstor.org/stable/1175860>
- Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Stein, M.K., Grover, B.W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, 33(2), 455–488. <https://doi.org/10.3102/00028312033002455>
- Stone, G.E. (1995, January). *Objective standard-setting*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Stone, G.E. (2001). Objective standard-setting: Or truth in advertising. *Journal of Applied Measurement*, 2(2), 187–201.
- Taylor, N. (2018). Teacher knowledge in South Africa. In N. Spaull & J. Jansen (Eds.), *South African schooling: The enigma of inequality*. Johannesburg: Springer. Retrieved from <https://www.springer.com/gp/book/9783030188108>
- Tiratira, N.L. (2009). Cutoff scores: The basic Angoff method and the Item Response Theory Method. *The International Journal of Educational and Psychological Assessment*, 1(1), 27–35.
- Van Geel, M., Keuning, T., Frèrejean, J., Dolmans, D., Van Merriënboer, J., & Visscher, A.J. (2019). Capturing the complexity of differentiated instruction. *School Effectiveness and School Improvement*, 30(1), 51–67. <https://doi.org/10.1080/09243453.2018.1539013>
- Venkat, H., Bowie, L. & Alex, J. (2017, October). *The design of a common diagnostic mathematics assessments for first year B.Ed. students*. Paper presented at the South African Education Research Association Conference, Port Elizabeth.
- Venkat, H., & Spaull, N. (2015). What do we know about primary teachers' mathematical content knowledge in South Africa? An analysis of SACMEQ 2007. *International Journal of Educational Development*, 41, 121–130. Retrieved from <http://hdl.handle.net/10539/17772>
- Wright, B.D., & Stone, M.H. (1979). *Best test design: Rasch measurement*. Chicago, IL: Mesa Press. Retrieved from <https://research.acer.edu.au/measurement/1/>
- Zieky, M., & Perie, M. (2006). *A primer on setting cut-scores on tests of educational achievement*. Educational Testing Service. Retrieved from [https://www.ets.org/Media/Research/pdf/Cut\\_Scores\\_Primer.pdf](https://www.ets.org/Media/Research/pdf/Cut_Scores_Primer.pdf)