International Journal of Instruction
e-ISSN: 1308-1470 ● www.e-iji.net

July 2023 ● Vol.16, No.3
p-ISSN: 1694-609X
pp. 1075-1094

Article submission code:
20220902235806

Received: 02/09/2022
Revision: 24/02/2023

Accepted: 20/03/2023
OnlineFirst: 01/06/2023

Trustworthiness of Teacher Assessment and Decision-Making: Reframing the Consistency and Accuracy Measures

Dennis Alonzo
School of Education, University of New South Wales SYDNEY, Australia. *d.alonzo@unsw.edu.au*
Steven Teng
School of Education, University of New South Wales SYDNEY, Australia. *s.teng@unswalumni.com*

The quality of assessment tools and the inferences drawn from the results to inform decisions in the classroom are usually measured using reliability and validity. These psychometric principles have been criticised for their inapplicability to classroom assessment, resulting in a parallel set of ‘classroometric’ principles. However, the use of two parallel principles widens the perceived dichotomy between formative and summative assessments. To overcome this dichotomy and ensure consistency of teachers’ decision-making, the concept of trustworthiness, drawn from qualitative research methodology, is increasingly being adopted, but it is under-theorised. We used a scoping technique to explore how this concept has been used in the assessment literature since it was first introduced in 1993. We accessed journal articles from four databases using combinations of search terms, resulting to 1,872 papers. Upon removal of duplicates and application of exclusion criteria, 27 papers remain relevant for full analysis. Our analysis expands Lincoln and Guba’s (1985) four criteria of qualitative research (credibility, transferability, dependability and confirmability) to include authenticity, rigour, fairness, equity, consistency, defensibility, accuracy, and adequacy and appropriateness of data. We develop a framework and a working definition for understanding trustworthiness in the context of assessment.

Keywords: trustworthiness, assessment, reliability, validity, ‘classroometric’ principles

INTRODUCTION

This paper builds on a notion of assessment to support student learning and teacher teaching that draws on both formative assessment (FA) and summative assessment (SA) strategies to support student learning, and argues that the traditional quality measures (psychometrics and “classroometric”) applied to SA and to FA do not capture the current conceptualisation of effective assessment practices, and thus the need for the concept of trustworthiness as an alternative measure. The rigour of teachers’ decision-making in the classroom is dependent on these quality measures alongside social and

Citation: Alonzo, D., & Teng, S. (2023). Trustworthiness of teacher assessment and decision-making: Reframing the consistency and accuracy measures. *International Journal of Instruction*, *16*(3), 1075-1094. https://doi.org/10.29333/iji.2023.16357a

statistical moderation, and the absence of such compromises the integrity of classroom assessments and teachers' decisions (Brown, 2019).

The quality of assessment tools and the inferences drawn from the assessment results used by teachers to inform decisions in the classroom are important considerations in evaluating students' learning, with the concepts of reliability and validity commonly used to measure assessment consistency and accuracy, respectively. Reliability measures the assessment tools' internal consistency or their ability to produce the same results across time while validity measures their accuracy whether the results really represent the outcomes measured or their ability to give valid inferences about student learning (Tabachnick & Fidell, 2007). These properties are critical aspects of assessment because any measurement error would significantly affect results (Field, 2009) and assessment results would not reflect students' true learning, and hence, teachers' decision-making is compromised. However, reliability and validity are often associated with the psychometric principles of assessment and are tied to SA, including formal tests and high-stake examinations, which are predominantly used for accountability. This is problematic as SA is not the only type of assessment that teachers use to support student learning. In fact, if SA is improperly used, it has negative consequences for learning and teaching (Harlen, 2007). Inside the classroom, teachers use assessments ranging from in-class FA to a more formal SA (Davison, 2007). For many such assessments, the concepts of reliability and validity cannot be applied because assessment activities such as self-assessment, peer assessment, feedback, and questioning are more informal and ongoing (Alt & Raichel, 2022), with no statistical metrics to establish their psychometric properties. Gu (2021) offers an argument-based framework to argue for the validity of formative assessment as an alternative approach. van der Vleuten and Schuwirth (2005) argue that the overreliance of assessment on psychometric issues limits the utility of other assessment strategies.

To ensure consistency in FA and to address the limitations of reliability and validity as FA cannot provide these metrics, Brookhart (2003), Moss (2003) and Smith (2003) developed the 'classroometric' principles of assessment. According to Brookhart, "the classroom assessment environment, the integration of assessment and instruction, and the pervasive formative purpose of classroom assessment" (p.8) need to be accounted for when discussing the consistency and accuracy issues of assessment. The focus of classroometric principles is on ensuring high-quality assessments that provide rigorous information about students' learning and immediately become part of the learning environment.

The classroometric principles offer an alternative reliability measure known as the sufficiency of information (Smith, 2003). Consistency of assessment information is drawn from multiple sources and integrating them to form a coherent picture of student learning. This conceptualisation overcomes the irrelevant factors in psychometrics that contribute to the inconsistency of information by providing sufficient information for decision making. The goal of consistency in classroometric principles is to have stable information derived from multiple sources about the gap between students' work and ideal work. As FA involves anecdotal records, observations, interviews, classroom tests and many other informal assessment activities, teachers need to reflect on all information available and decide to what extent individual students have achieved their expected learning outcomes.

However, in the current conceptualisation of assessment where every information is used to support student learning and teaching (ARG, 1999, 2002; Baird et al., 2017), the psychometrics and 'classroometric' principles are not enough to account for the various factors that contribute to the consistency and accuracy measures of assessment and assessment data (Alonzo, 2016). These measures are not inherent characteristics of assessments and teachers' decision-making processes, but rather influenced by many personal and contextual factors (Alonzo et al., 2021; Taylor, 2013; van der Vleuten & Schuwirth, 2005). The "classroometric" principles do not account for teachers' knowledge of theory, technical skills, principles and concepts, pedagogy, sociocultural values, local practices, personal beliefs and attitudes as necessary requisites for effective assessment and decision-making (Taylor, 2003). In addition, focusing on using "classroometric" principles will not ensure consistency of teachers' decision-making as FA is heavily critiqued for their inability to provide sufficient evidence of student learning (Brown, 2019).

The use of two parallel principles appears to have had the opposite effect to that intended, widening the perceived dichotomy between formative and summative assessments. This dichotomy is irrelevant to the current conceptualisation of assessment to support student learning, where formative and summative assessment strategies are inherently interlinked. To address the issues above, this paper aims to argue for the need of an alternative consistency measures. We explored how the concept of trustworthiness, drawn from qualitative research methodology, is adopted in assessment. This concept is under-theorised in assessment, and thus, this paper is an attempt to provide a coherent understanding how it can be used to ensure consistency of teacher assessment and decision-making process.

Literature Review

Limitations of Psychometric and Classroometric Principles

As discussed above, both psychometrics and 'classroometric' principles have been used extensively to account for the consistencies of assessments. However, they are not aligned with the current conceptualisation of assessment. This section discusses in detail the limitations of both principles.

Although the concepts of reliability and validity are critically important in the field of assessment, there are limitations in terms of their applicability to classroom assessment, especially more informal and contingent formative assessment. The following issues highlight the limitations of these two concepts.

First, there is a growing argument that strict adherence to these measurement principles divides theories and concepts from practice. For example, Brookhart (2003) argues that authentic assessment does not draw much on traditional concepts of measurement. Instead, in actual assessment and teaching practices, the emphasis shifts from measurement principles to the quality of information, considering the context-dependence of assessment, the use of assessment information for teaching, and the formative and summative functions of any assessment. This argument is supported by Smith (2003), who called for a reconsideration of reliability to account for the context under which assessment occurs. He argues that the different forms of reliability measures negate the principles of learning. For example, the test-retest or parallel forms of reliabilities, founded on the concept of stability of students' performance across different times of testing, are impractical for classroom assessment as each teacher works hard to bring about significant change to the performance of students in the least possible time.

Along the same line of argument, the coefficient alpha, which is tied to the concept of score variance, is an unreasonable measure if invoked for classroom assessment. In psychometrics, any items that all students get right or wrong do not affect test variance, and hence are considered useless items. However, in classroom assessment, these items are important for teachers to determine which learning objectives have been achieved by students and/or which learning objectives need further exploration to enhance student learning. Because of the inappropriateness of psychometric measures of reliability, Smith (2003) proposes an alternative conceptualisation of reliability as sufficiency of information to make robust decisions about learning and teaching. This conceptualisation of reliability is contextualised in the nature of classroom learning and teaching, which accounts for "the multidimensionality of the underlying assessment and does not require a rank order of the students" (p. 31).

Similarly, the traditional view of validity, which has evolved from Thorndike (1918) criterion-based model of validity to Messick (1989) content-based validity model, including Cronbach and Meehl (1955) construct model of validity, requires reconsideration. In this regard, Moss (2003) reconceptualises validity in the context of classroom assessment using an interpretive approach. She challenges the assumptions of traditional validity by reflecting on actual classroom practices. First, contrary to the view that assessment is a distinct episode in learning and teaching, in the actual classroom setting, assessment is an integral part of the learning and teaching processes. Assessment forms the network that binds all other classroom activities to support student learning. Second, the concept that validity requires the appropriateness of interpretation of student learning based on the assessment results is too limiting or too encompassing. According to Moss, in actual practice, when teachers are concerned about student progress, there is no need to have a fixed interpretation of student competence. Rather, teachers need to make trustworthy decisions regularly using assessment information to support student learning and monitor their decisions' learning consequences. Various assessment results should inform these decisions, as a single assessment cannot provide sufficient information. Thus, when considering the validity of individual assessment practices, one should look into how each assessment "fits with the other assessment practices, in progression, to support (and illuminate) learning" (Moss, 2003, p. 16). Third, the consideration that the individual student is the unit of analysis excludes the role of classroom context in assessment and learning. Moss elaborates the argument of Mehan (1998) on using the social situation as the unit of analysis, emphasizing that learning and teaching decisions should be based on evidence derived from the analysis of the interactions of all the classroom activities. Fourth, the assumption that interpretations are based on combining all judgments from assessment results will not provide any convincing evidence in situations where the aggregation is impossible or undesirable. Moss points out that drawing the right interpretations about student learning is an iterative process that involves repeated measurement of student learning. This requires teachers to test the pieces of evidence by gathering information from various sources until these pieces of evidence can be formed into a coherent picture of student achievement. Fifth, any assessment practice has a consequence contrary to the widely accepted concept that consequence matters only if the source of a particular consequence can be traced to construct underrepresentation or construct irrelevant variance. This last view was further elaborated by McNamara and Roever (2006) by studying in detail the social consequences of assessment.

Other researchers have supported Moss' arguments described. Shepard (2001) argued nearly fifteen years ago that in a less standardized test, there is a need for new methods of analysing and interpreting students' responses. Kane (2001) proposed an argument-based approach to validity in which he "suggests that the proposed interpretation be specified in terms of a network and inferences and assumptions, that these inferences and assumptions be evaluated using all available evidence, and that the plausible alternate interpretations be considered" (p. 339-340). In a more detailed way, Killen (2003) looked into the processes that help improve the assessment's validity He argues that all classroom activities influence the appropriateness and usefulness of teachers' judgments, from developing learning outcomes to providing learning opportunities that help students achieve targets. Killen emphasises the necessity to ensure the coherence of all the learning, assessment, and teaching activities. The interplay and the direct focus of all these activities on student learning ensure the trustworthiness of the teachers' decisions on student learning. Clearly, as each learning and assessment practice evolves in various classroom contexts, there is a need to reframe the concepts of reliability and validity to the context of actual and authentic assessment practice.

The 'classroometric' principles are also limited in their capacity to account for the consistency measure of the current conceptualisation of assessment. First, if we continue to use both psychometrics and 'classroometric' principles separately, it will continue to widen the dichotomy between FA and SA, a distinction that has been refuted by several authors (Black, 2017). Second, 'classroometric' principles, like the psychometrics, do not account for other factors that contribute to the consistency of assessment. As documented in the literature, there are contextual (Daugherty et al., 2011), personal, both teachers and students (Black, 2015), political (Davison, 2013), and other factors that contribute to the quality of assessment and the inferences drawn. Third, teachers draw from their professional judgment to make important decisions for individual students in the classroom. They use both FA and SA results and often conflate psychometrics and 'classroometric' principles (Lau, 2016).

Given the limitations discussed above, and the current conceptualisation of assessment, psychometrics and 'classroometric' principles are insufficient to account the various factors that contribute to the consistency of assessment. To account for the limitations of these principles, an alternative concept of assessment trustworthiness has been developed (Alonzo, 2016; Davison, 2007 & Gipps, 1994), but it is undertheorised. This literature review aims to explore how the concept of trustworthiness is being used in the assessment literature through a review of scholarly work. We aim to clearly define the concept based on the extant literature and develop a framework for understanding and ensuring the trustworthiness of assessment.

The Concept of Trustworthiness

To account for the limitations of psychometrics and 'classroometric' principles, trustworthiness has been introduced as an alternative concept, but it is undertheorised in the field of assessment. The concept of trustworthiness has been extensively used in the qualitative research literature to ensure rigour and legitimacy parallel to reliability and validity measures in quantitative research. Trustworthiness was first proposed as a concept by Lincoln and Guba (1994), with credibility, transferability, dependability and confirmability as its subs-constructs. The utility of trustworthiness and its subconstructs has been extensively used and explored in qualitative research (Loh, 2013, Nowell et al., 2017).

In the field of assessment, the concept of trustworthiness was first used by Hipps in 1993 in his report, arguing for new methods to judge alternative assessment quality. He suggests that Lincoln and Guba's (1985) criteria for trustworthiness better represent the constructivist paradigm of authentic assessment, where the perception that knowledge is created through experiences has no perfect criteria, thus requiring scrutiny that can measure each student's development, instead of the traditional measurements of reliability and validity. He argues that the four criteria of trustworthiness are parallel with validity, generalizability, reliability, and objectivity. He adds authenticity and fairness of assessment to account for the views and perception of stakeholders about assessment. A year later, Gipps (1994) published a book and outlines the qualities of a trustworthy assessment using Lincoln and Guba's criteria. She puts forward that:

> credibility comes from prolonged engagement and persistent observation, i.e., regular ongoing assessment in the classroom, and including parents and in the dialogue about pupil performance. Transferability could replace the notion of generalizability: since performance is context bound the assessor must specify the context in which a particular achievement was demonstrated… Dependability replaces traditional reliability it is related to the process of assessment and the judgments made which must be open to scrutiny…Authenticity is to do with the extent to which the relevant constructs (and this means all stakeholders' constructs) are fairly and adequately covered in the assessment (p.168).

Later on, in 2003, Webb et al., explicitly applied trustworthiness and demonstrated how its subconstructs provide a more rigorous assessment of student portfolio. They added more subconstructs, including adequacy and appropriateness of assessment data, to fully capture the consistency measures of assessment. The limitations in any criterion of assessment quality were highlighted by van der Vleuten and Schuwirth in 2005 and proposed optimising assessment at a program level. In their more recent paper, they have used the subconstructs of trustworthiness as follows: “the dependability and credibility of the overall decision relies on the combination of the emanating information and the rigour of the supporting organisational processes’ (van der Vleuten et al., 2015, p.642). Other authors used quality-related measures of assessment, including defensibility (Bacon et al., 2015), rigour (Blackburn, 2019), fairness (Harlen, 2005), equity (Scott et al., 2014), consistency (Connolly et al., 2012), and accuracy (Harlen, 2005).

METHOD

A scoping review methodology was employed to explore how trustworthiness is used in assessment. This involved generating research questions and charting data to synthesise concepts and identify gaps in the literature, following the five steps outlined by Arksey and O’Malley (2005). Our team of three researchers worked collaboratively at each stage. Our interactions and dialogue helped us to engage deeper in our literature review (Andrews, 2005) and developed our greater understanding of trustworthiness as used in the literature.

Stage 1: Identifying the research question. Based on the aim of this paper, the research question was developed to guide the reading of the literature to which this review sought to answer: *How is trustworthiness used in the current assessment literature?*

Stage 2: Identifying relevant studies. The literature was sourced from the following online databases: Proquest Education, A+ Education, Web of Science and ERIC. The keyword ‘assessment’ was searched in conjunction with ‘trustworthiness’. We replaced the ‘trustworthiness’ with subconstructs associated with it such as: ‘credibility’, ‘transferability’, ‘dependability’, “defensibility’, ‘confirmability’, ‘authenticity’, ‘rigour’, ‘fairness’, ‘equity’, consistency’, ‘adequacy’ and ‘accuracy’. We included only journal articles that had been peer-reviewed without a strict boundary on the year of publication to trace the introduction and development of the concept. We included literature from school settings, vocational education and higher education. Literature that did not address assessment in the classroom context or relate to education was not included in the review. We also complemented our search in databases by following citations flows and searching the refence lists of relevant literature.

Stage 3: Study selection. Many of the search terms identified a larger number of results. We discussed which literature were most relevant by applying a set of inclusion and exclusion criteria (Arksey & O’Malley, 2005). We read the abstract and keywords and decided if the criteria we set were met. We included peer-reviewed journal articles, research-based and specialist contributions (Tight, 2012). After removing duplications and publications that are not directly related to the research question, 27 remained and were included in the review. Figure 1 summarises the approach undertaken for Stages 2 and 3.

Stage 4: Data charting. With the number of articles finalised, each was annotated, and the terminology (trustworthiness or its related sub-constructs) used was identified. We worked independently and charted the key arguments in how the term was defined or used by the authors. We used the following categories:

- Details (author, year of publication, topic)
- Concept used (trustworthiness, rigour, authenticity, fairness, consistency)
- Key argument (definition/how it is used)

This allowed us to analyse each reported area of research focus and findings and group the articles into themes. After we individually coded the literature, we discussed our annotations and integrated our annotations to best present each literature.

Stage 5: Collating, summarising and reporting. We examined our annotations to address our research questions. We summarised the findings in relation to the themes (Arksey & O'Malley, 2005), using the sub-constructs of trustworthiness. We negotiated for codes that we did not agree by re-reading the journal article. This approach allowed us to present how trustworthiness was used in assessment in a narrative form. We compared how each author used the concept and reported the similarities and differences. As this is a scoping review, we did not intend to provide a weight of evidence for each sub-construct of trustworthiness.

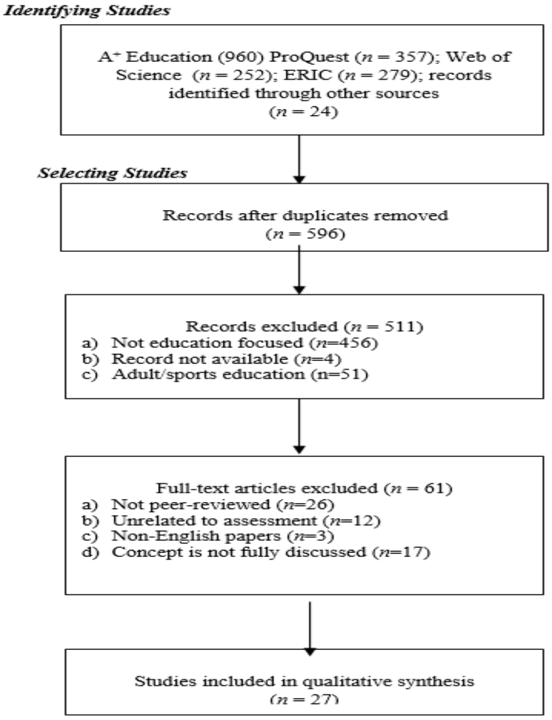
Identifying Studies
A⁺ Education (960) ProQuest (n = 357); Web of Science (n = 252); ERIC (n = 279); records identified through other sources (n = 24)
Selecting Studies
Records after duplicates removed (n = 596)
Records excluded (n = 511)
a) Not education focused (n=456)
b) Record not available (n=4)
c) Adult/sports education (n=51)
Full-text articles excluded (n = 61)
a) Not peer-reviewed (n=26)
b) Unrelated to assessment (n=12)
c) Non-English papers (n=3)
d) Concept is not fully discussed (n=17)
Studies included in qualitative synthesis (n = 27)

Figure 1
Study selection flow diagram

FINDINGS

The results are presented following the research question:

How is trustworthiness used in the current assessment literature?

As mentioned earlier, the concept of trustworthiness in assessment is under-theorised. It is not surprising that only 27 peer-reviewed papers have focused on this construct or its associated concepts.

Explicit use of trustworthiness

Our scoping study shows that there are only four peer-reviewed article that explicitly use the term trustworthiness despite being used in reports, theses and professional development resources. Meisels et al. (2001) argue that to achieve trustworthy assessment decisions, teachers need to explore students' background, including levels of ability and use this information to support their learning. Assessment information should come from multiple sources, and any inconsistencies with the information would require obtaining more information to make decisions more trustworthy and defensible (van der Vleuten & Schuwirth, 2005). According to van der Vleuten et al., (2015), "trustworthiness of assessment and decision-making requires many data points of rich information, that is, resting on broad sampling across contexts, methods and assessors (p.643)." The consistency and accuracy of assessment and decision-making largely impact the effectiveness of students' learning (Buhagiar, 2007).

Carless (2009) used the term 'trust' to refer to "the confidence one has in the likelihood of others (management, administration, colleagues, students) acting responsibly in respect of sound principles, practices or behaviours in assessment" (p.81). He argues that trustworthy assessments should favour learning-orientated assessment practices to maximise learning through peer and self-assessment and innovative activities and assessments, promoting productive learning rather than awarding marks.

The four criteria of trustworthiness

Apart from trustworthiness and trust, the four criteria of trustworthiness reported by Lincoln and Guba (1985) and Gipps (1995) are also used.

Credibility is used in the literature to discuss the importance of how assessment data sources are identified and described accurately, including the process used to assess student learning (Webb et al., 2003). It is also used to position the importance of why teachers need to have a rich understanding of the standards and why assessment approaches need to be dynamic to account for emerging knowledge and skills.

Transferability of assessment activities involves adapting them for specific context contributes to trustworthiness (Billing & Thomas, 2000). There are cultural, structural, political and technical issues influencing the transferability of assessment.

There are three pieces of literature that use the concept of dependability. This refers to how the assessment results can provide meaningful inferences for teachers to make significant decisions to support students. Wiliam (1993) draws on the concepts of reliability and validity and applies it to all range of assessment activities, while Webb et al. (2003) discuss the processes required to make the inferences dependable. Teachers need to verify evidence, engage in moderation activities and adhere to external and internal quality assurance processes. Similarly, the clarity of learning outcomes, the use of explicit grading criteria, and the curriculum's transparency contribute to dependability (Harlen, 2005).

Many articles use the term authentic assessment but only two use the term authenticity that characterises assessment. Villarroel et al. (2018) used it to highlight assessment design to promote learning while Palmer (2004) used it to argue that assessment tasks should reflect and develop knowledge and skills that can be applied in real-world contexts.

Other related constructs

Apart from these four criteria of trustworthiness, eight other concepts were reported in the reviewed articles that contribute to consistency and accuracy measures of teachers' assessment and decision making.

The concept of rigour in assessment was discussed extensively how it is linked to trustworthiness in six articles. Medaille et al. (2019) suggest to carefully plan the assessment tasks to reflect the specificity of individual students' unique level of development and competency. Assessments should promote high levels of critical thinking required in the real world, aligning with the notion of authentic assessment, where students should be assessed against a continuum of the learning outcomes. In addition, Webb et al. (2003) discuss various processes that contribute to the rigour of assessment. Monitoring assessment implementation is needed to ensure rigour and quality and to ensure that teachers provide regular and quality feedback, double mark and moderate for increased internal validity, and adhere to external quality assurance schemes. Colbert et al. (2012) add to this by suggesting that sustainable assessment culture contributes to the rigour of assessment, and this can be maintained through leadership in learning, which can ensure teachers align curriculum, pedagogy and assessment, and design quality assessment tasks with clear standards, expectations, evidenced-based judgements and moderation.

Fairness is discussed in ten articles as a construct that allows students to be assessed without any bias. Stobart (2005) argues that assessments can never be entirely fair due to cultural diversity and equality complexities. However, fairness can be achieved through dynamic interplay among teaching, learning, classroom interactions and assessment elements with a focus on supporting individual students' learning from diverse backgrounds (Rasooli et al., 2018). In addition, it can be achieved by using learner-centred assessments such as portfolios, projects and collaborative assessments as they are developed over time, having sought negotiation from a mixture of peers and teachers and integrated different feedback and perspectives to produce a quality work (Flores et al., 2014). Yung and Yung (2001) emphasise that students need to be assessed on a fair basis without jeopardising their chances to learn while they are being assessed. Harlen (2005) and Alm and Colnerud (2015) argue that fairness requires teachers to be aware of the sources of bias in their assessments and avoid the influence of irrelevant factors. Moreover, fairness can be achieved by using alternative assessment practices alongside the need for teachers' pedagogic practice to cater for increased diversity and working with other stakeholders to support students (Klenowski, 2014). Furthermore, the explicit use of criteria, frequency of feedback and communication of the results to students (Pepper & Pathak, 2008) and supporting students to achieve the assessment outcomes and addressing their expectations contribute to fairness (Lantolf & Poehner, 2013).

Four articles expanded fairness to include addressing equity. Driver (2019), Scott et al. (2014), and Siegel (2007) put forward that assessment must be differentiated to accommodate students' ability, race, ethnicity, culture, language, and socio- economic status. Equity in assessment involves evaluating class content, adapting and diversifying tests and other assessments, even taking into account students' effort and attitudes (Murillo & Hidalgo, 2017) and the role of context and construct (Gipps, 1995).

Another concept that accounts for trustworthiness is consistency which was discussed in three articles. Teachers need to improve consistency of assessment and judgements on learning to align and report report student work against curriculum standards (Meiers et al., 2007). Methods to achieve this include constant moderation and review of student work to calibrate against performance standards. Context was identified as an important influential factor in teachers' judgements; teachers' assessment beliefs, attitudes and practices impact on their perceptions of the value of moderation practice and the extent to which consistency can be achieved (Connolly et al., 2012). Consistency in assessment is also achieved through school leadership in establishing a 'assessment culture' in which assessment is discussed constructively and positively and not seen as a necessary chore (Harlen, 2005).

The other concept used in the literature is the accuracy of assessment. This means that assessment results should reflect how individual students have achieved their learning goals (Harlen, 2005). The decisions made by teachers should be based on the adequacy and appropriateness of data, which is best done with the ongoing collection of evidence of student learning in a variety of settings. The aim is to compare these pieces of evidence and establish a holistic picture of how students are meeting the standards (Webb et al., 2003). Additional concept, confirmability, was used by Webb et al., (2003) to argue for the need of accurate assessment data management. Any assessment data should be linked to their sources so that teachers can verify evidence during moderation processes and check inferences drawn. A related concept, defensibility, is used by Bacon et al. (2015) to argue that data should support teachers' decisions.

Overall, what is highlighted in this section is that trustworthiness and its related constructs are used to refer to various characteristics of assessment, including its quality, design, processes, implementation, and quality assurance mechanisms implemented by schools and teachers. It also includes assessment literacy including views, beliefs, knowledge and skills of principals, teachers, and students operating within the school context and policy constraints.

DISCUSSION

Based on the literature discussed above, although the term trustworthiness is under-theorised, it has been used and started to gain significant attention as more studies have explored this concept and its sub-constructs. It was shown that the trustworthiness of teacher assessment and decision-making incorporates various consistency and accuracy measures, expanding the psychometrics and 'classroometric' principles of assessment. Based on the answers to Research Question 1, trustworthiness in assessment and decision-making expands Lincoln and Guba's (1985) four criteria of qualitative research (credibility, transferability, dependability and confirmability) to include authenticity, rigour, fairness, equity, consistency, defensibility, accuracy, and adequacy and appropriateness of data. All these subconstructs of trustworthiness have to do with the quality of the assessment plan and design, the ability of teachers for ensuring accurate process, including implementing assessment, gathering and analysing assessment data to inform decision making, the involvement of students and other stakeholders, within the backdrop of school context and policy. In other words, ensuring trustworthiness of assessment requires ensuring the quality of the entire assessment approach, from planning to using assessment data for decision-making.

Based on our findings, we identified the key elements that contribute to the trustworthiness of assessment and teacher decision-making. These elements are reflected in the framework we developed (Figure 2) for understanding the trustworthiness of assessment, and to guide teachers on how to ensure the trustworthiness of their assessment practices particularly in their decision-making process.

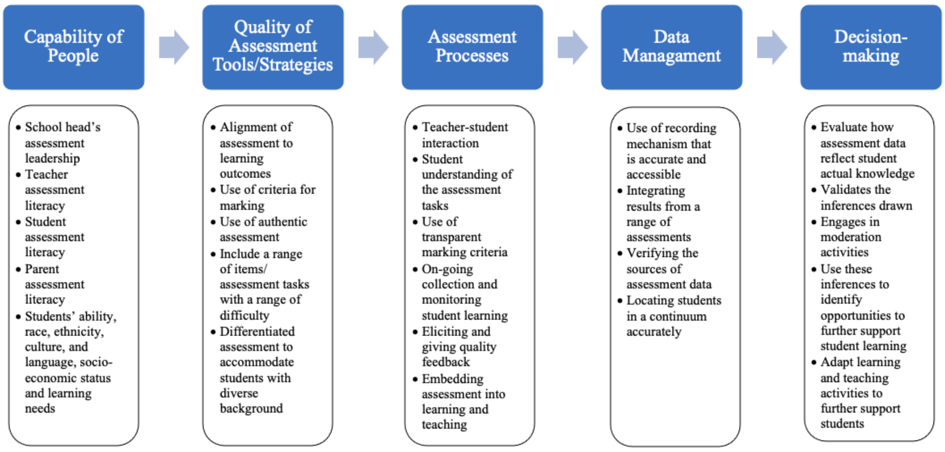
Capability of People
Quality of Assessment Tools/Strategies
Assessment Processes
Data Managament
Decision-making
• School head's assessment leadership
• Teacher assessment literacy
• Student assessment literacy
• Parent assessment literacy
• Students' ability, race, ethnicity, culture, and language, socio-economic status and learning needs
• Alignment of assessment to learning outcomes
• Use of criteria for marking
• Use of authentic assessment
• Include a range of items/ assessment tasks with a range of difficulty
• Differentiated assessment to accommodate students with diverse background
• Teacher-student interaction
• Student understanding of the assessment tasks
• Use of transparent marking criteria
• On-going collection and monitoring student learning
• Eliciting and giving quality feedback
• Embedding assessment into learning and teaching
• Use of recording mechanism that is accurate and accessible
• Integrating results from a range of assessments
• Verifying the sources of assessment data
• Locating students in a continuum accurately
• Evaluate how assessment data reflect student actual knowledge
• Validates the inferences drawn
• Engages in moderation activities
• Use these inferences to identify opportunities to further support student learning
• Adapt learning and teaching activities to further support students

Figure 2
Framework for ensuring trustworthiness of teacher assessment and decision-making

Although the framework follows a step by step process, each step is not isolated from the others. Rather, they are all interconnected and support each other. The key elements are as follows:

1. The capability of the people involved. This includes teachers' assessment literacy (Alm & Colnerud, 2015; Klenowski, 2014) (Siegel, 2007), students' assessment literacy, including their beliefs and views of the role of assessment in their learning, to have a deeper understanding of the purpose to effectively engage in any assessment activity (Flores et al., 2015), parents' involvement in assessment activities and other stakeholders' participation in assessment (Klenowski, 2014). Also, the school leadership team contributes trustworthiness of assessment through the strategic development of assessment culture (Alm & Colnerud, 2015; Harlen, 2005).
2. The quality of assessment design, tools and strategies used. This includes the authenticity of assessment (Villarroel et al., 2018) - how it reflects the skills developed and measured (Palmer, 2004), ensures equity (Driver, 2019; Scott et al., 2014) by adapting assessment to cater students' diverse backgrounds (Murillo & Hidalgo, 2017) (Siegel, 2007), avoids sources of bias (Harlen, 2005) (Alm & Colnerud, 2015), and guarantees students' success (Yung & Yung, 2001).
3. The assessment processes used to optimise the use of assessment to support student learning. This includes the pedagogical approaches used by teachers in embedding assessment in learning and teaching (Colbert et al., 2012; Rasooli et al., 2018), use of explicit criteria and standards (Pepper & Pathak, 2008), modification of assessment activities to account student diverse backgrounds, and interactions of teachers and students (Klenowski, 2014).
4. The data management system used by teachers and schools in general. This include using a recording mechanism that is accurate and accessible (Webb et al., 2003), integrating the range of assessment data from various sources (Bacon et al., 2015; Klenowski, 2014), and locating students in a continuum accurately (Medaille et al., 2019).
5. The decision-making processes. This include engaging in moderation activity to make consistent judgement (Colbert et al., 2012), validating the inference drawn from the data (Webb et al., 2003; Wiliam, 1993), using these inferences to identify opportunities to further support student learning (Lantolf & Poehner, 2013), evaluate how assessment data reflect student actual knowledge (Harlen, 2005), and adapting learning and teaching activities to further support students (Murillo & Hidalgo, 2017).
6. The impact of contextual, cultural and personal factors. Teachers need to consider the effect of students' background, learning needs and cultural orientation when engaging in assessment and how they demonstrate their learning (Driver, 2019). Teachers' assessment practices should be strongly underpinned by fairness, equity, access and equality (Cherry et al., 2003; Driver, 2019; Murillo & Hidalgo, 2017; Rasooli et al., 2018).
7. The influences of assessment policies. They shape the assessment culture in schools (Harlen, 2005). Also, teachers need to consider how cultural, structural, political and technical issues influence the effectiveness of assessment (Billing & Thomas, 2000).

This definition of trustworthiness in assessment has a significant theoretical implication in addressing the issues of inappropriateness of psychometric 'classroometric' principles to account for the continuum of assessment practices, widening the perceived dichotomy between formative and summative assessments. Using trustworthiness to argue for the accuracy and consistency of teacher assessment practices and decision-making, would support teachers to use a range of assessment strategies and to integrate all types of assessment data and use their professional judgment to determine where "students are in their learning, where they need to go, and how best to get there" (ARG, 2002, p. 2). In addition, using trustworthiness will avoid the debate around the reliability and validity of formative assessment, thus emphasising the philosophical nature of assessment as a continuum of practice (Black, 2017; Davison, 2007). This will reinforce the complementary roles of FA and SA in supporting student learning (Lau, 2016).

CONCLUSION AND IMPLICATIONS

Our literature review builds on the notion of assessment to support student learning and teacher teaching that integrates both FA and SA and argues that the current consistency and accuracy measures widely used with FA (classroometric principles) and SA (psychometric principles) do not capture the present conceptualisation of effective assessment practices. Thus, introducing the concept of trustworthiness as an alternative measure. From our scoping study, it can be seen that the concept of trustworthiness and its subconstructs are gaining significant attention in the field of assessment. Our findings have implications in advancing the theorisation of trustworthiness. First, trustworthiness in assessment is not only limited to its original four criteria (credibility, transferability, dependability and confirmability), but it includes authenticity, rigour, fairness, equity, consistency, defensibility, accuracy, and adequacy and appropriateness of data. All these subconstructs enhance the accuracy and consistency measures of any assessments to optimise their functions for supporting student learning and teacher teaching. Second, trustworthiness has to do with the quality of assessment plan and design, the ability of teachers for accurate process including gathering and analysing assessment data to inform decision-making, the beliefs, views, knowledge, skill, and active engagement of students and other stakeholders, within the backdrop of school context and policy.

There are some limitations of our study that we have identified. Although our paper provides the first coherent understanding of how trustworthiness is used and reported in the literature after searching through databases, we only included peer-reviewed journal articles. This exclusion criterion may have limited our literature search. We did not review printed books and documents from the government and international agencies about principal data literacy. These other publications can be considered in future systematic literature reviews.

Our findings have implications for future research. We need empirical studies that will support our proposed conceptualisation and the viability of the framework for enhancing the trustworthiness of teachers' assessment practices and decision making. Studies are also needed to explore other factors and processes that contribute to the trustworthiness of assessment and teacher decision-making These factors and processes will lead to identifying specific indicators of trustworthiness that can be used for empirical studies to establish its dimensionality.

REFERENCES

Alm, F., & Colnerud, G. (2015). Teachers' experiences of unfair grading. *Educational Assessment, 20*(2), 132-150. https://doi.org/10.1080/10627197.2015.1028620

Alonzo, D. (2016). *Development and application of a teacher assessment for learning (AfL) literacy tool.* University of New South Wales]. Sydney. http://unsworks.unsw.edu.au/fapi/datastream/unsworks:38345/SOURCE02?view=true

Alonzo, D., Labad, V., Bejano, J., & Guerra, F. (2021). The policy-driven dimensions of teacher beliefs about assessment. *Australian Journal of Teacher Education, 46*(3). https://ro.ecu.edu.au/cgi/viewcontent.cgi?article=4761&context=ajte

Alt, D., & Raichel, N. (2022). Problem-based learning, self- and peer assessment in higher education: towards advancing lifelong learning skills. *Research Papers in Education, 37*(3), 370-394. doi:10.1080/02671522.2020.1849371

Andrews, R. (2005). The place of systematic reviews in education research. *British Journal of Educational Studies, 53*(4), 399-416. http://www.jstor.org/stable/3699275

ARG. (1999). *Assessment for Learning: Beyond the Black Box*. http://hdl.handle.net/2428/4621

ARG. (2002). *Assessment for learning: 10 principles*.

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19-32. https://doi.org/10.1080/1364557032000119616

Bacon, R., Williams, L., Grealish, L., & Jamieson, M. (2015). Credible and defensible assessment of entry-level clinical competence: Insights from a modified Delphi study. *Focus on Health Professional Education, 16*(3), 57-72.

Baird, J.-A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy & Practice, 24*(3), 317-350. https://doi.org/10.1080/0969594X.2017.1319337

Billing, D., & Thomas, H. (2000). The international transferability of quality assessment systems for higher education: The Turkish experience. *Quality in Higher Education, 6*(1), 31-40. https://doi.org/10.1080/13538320050001054

Black, P. (2015). Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice, 22*(1), 161-177. https://doi.org/10.1080/0969594X.2014.999643

Black, P. (2017). Assessment in science education. In K. S. Taber & B. Akpan (Eds.), *Science Education: An International Course Companion* (pp. 295-309). SensePublishers. https://doi.org/10.1007/978-94-6300-749-8_22

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice, 25*(6), 551-575. https://doi.org/10.1080/0969594X.2018.1441807

Blackburn, B. R. (2019). *Ensuring rigour in a differentiated classroom* (Vol. 41). Australian Council for Educational Leaders. https://doi.org/10.3316/informit.513045935439865

Bloom, B. S., Hastings, J. T., & Madaus, G. (1971). *Handbook on formative and summative evaluation of student learning*. McGraw-Hill.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice, 22*(4), 5-12. https://doi.org/10.1111/j.1745-3992.2003.tb00139.x

Brown, G. T. L. (2019). Is assessment for learning really assessment? [Perspective]. *Frontiers in Education, 4*(64). https://doi.org/10.3389/feduc.2019.00064

Buhagiar, M. (2007). Classroom assessment within the alternative assessment paradigm: revisiting the territory. *The Curriculum Journal, 18*(1), 39-56.

Carless, D. (2009). Trust, distrust and their impact on assessment reform. *Assessment & Evaluation in Higher Education, 34*(1), 79-89. https://doi.org/10.1080/02602930801895786

Cherry, B., Ordóñez, L. D., & Gilliland, S. W. (2003). Grade expectations: the effects of expectations on fairness and satisfaction perceptions. *Journal of Behavioral Decision Making, 16*(5), 375-395. https://doi.org/https://doi.org/10.1002/bdm.452

Colbert, P., Wyatt-Smith, C., & Klenowski, V. (2012). A systems-level approach to building sustainable assessment cultures: Moderation, quality task design and dependability of judgement. *Policy Futures in Education, 10*(4), 386-401. https://doi.org/10.2304/pfie.2012.10.4.386

Connolly, S., Klenowski, V., & Wyatt-Smith, C. M. (2012). Moderation and consistency of teacher judgement: Teachers' views. *British Educational Research Journal, 38*(4), 593-614. https://doi.org/https://doi.org/10.1080/01411926.2011.569006

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.

Daugherty, R., Black, P., Ecclestone, K., James, M., & Newton, P. (2011). Assessment of Significant Learning Outcomes. In R. Berry & B. Adamson (Eds.), *Assessment Reform in Education: Policy and Practice* (pp. 165-183). Springer Netherlands. https://doi.org/10.1007/978-94-007-0729-0_12

Davison, C. (2007). Views from the chalkface: English language school based assessment in Hong Kong. *Language Assessment Quarterly, 4*(1), 37-68. https://doi.org/https://doi.org/10.1080/15434300701348359

Davison, C. (2013). Innovation in assessment: Common misconceptions and problems. In K. Hyland & L. Wong (Eds.), *Innovation and change in English language education* (pp. 263-275). Routledge.

Dorn, S. (2010). The political dilemmas of formative assessment. *Exceptional Children, 76*(3), 325-337. https://doi.org/10.1177/001440291007600305

Driver, M. K. (2019). Understanding equitable assessment: How preservice teachers make meaning of disAbility. *Journal of Multicultural Affairs, 4*(1).

Field, A. (2009). *Discovering statistics using SPSS*. Sage.

Flores, M. A., Veiga Simão, A. M., Barros, A., & Pereira, D. (2015). Perceptions of effectiveness, fairness and feedback of assessment methods: a study in higher education. *Studies in Higher Education, 40*(9), 1523-1534. https://doi.org/10.1080/03075079.2014.881348

Gipps, C. (1994). Beyond testin: Towards a theory of educational assessment. The Famer Press, taylor & Francis: London.

Gipps, C. (1995). What do we mean by equity in relation to assessment? *Assessment in Education: Principles, Policy & Practice, 2*(3), 271-281. https://doi.org/10.1080/0969595950020303

Gu, P. Y. (2021). An argument-based framework for validating formative assessment in the classroom. *Frontiers in Education, 6*. doi:10.3389/feduc.2021.605999

Harlen, W. (2005). Trusting teachers' judgement: research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education, 20*(3), 245-270. https://doi.org/10.1080/02671520500193744

Harlen, W. (2007). The impact of summative assessment on children, teaching, and the curriculum. In K. Möller, P. Hanke, C. Beinbrech, A. K. Hein, T. Kleickmann, & R. Schages (Eds.), *QualitÄt von Grundschulunterricht: entwickeln, erfassen und bewerten* (pp. 51-65). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90755-0_4

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge, Hoboken.

Herman, J. L., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2006). *The nature and impact of teachers' formative assessment practices*.

Hipps, J.A. (1993). Trustworthiness and authenticity: Alternate ways to judge authentic assessments. *ERIC Clearinghouse.* Washington DC.

Kane, M. T. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement, 38*(4), 319-342. https://doi.org/10.2307/1435453

Killen, R. (2003). Validity in outcomes-based assessment. *Perspectiv ein education, 21*(1), 1-14.

Klenowski, V. (2014). Towards fairer assessment. *The Australian Educational Researcher, 41*(4), 445-470. https://doi.org/10.1007/s13384-013-0132-x

Lantolf, J. P., & Poehner, M. E. (2013). The unfairness of equal treatment: objectivity in L2 testing and dynamic assessment. *Educational Research and Evaluation, 19*(2-3), 141-157. https://doi.org/10.1080/13803611.2013.767616

Lau, A. M. S. (2016). 'Formative good, summative bad?' – A review of the dichotomy in assessment literature. *Journal of Further and Higher Education, 40*(4), 509-525. https://doi.org/10.1080/0309877X.2014.984600

Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of family medicine and primary care, 4*(3), 324-327. https://doi.org/10.4103/2249-4863.161306

Loh. J. (2013). Inquiry into issues of trustworthiness and quality in narrative studies: A perspective. *The Qualitative Report, 18*(65), 1-15. http://www.nova.edu/ssss/QR/QR18/loh65.pdf

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry*. Sage. https://doi. org/10.1016/0147-1767(85)90062-8

Masters, G. (2015). Rethinking formative and summative assessment. *Teacher*. https://www.teachermagazine.com/au_en/articles/rethinking-formative-and-summative-assessment

McNamara, T., & Knoch, U. (2012). The Rasch wars: The emergence of Rasch measurement in language testing. *Language Testing*. https://doi.org/10.1177/0265532211430367

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Blackwell.

Medaille, A., Goldrup, S., & Abernathy, T. (2019). Assessing rigor in teacher education: Do NCTQ's guidelines measure up? *The Teacher Educator, 54*(1), 72-89. https://doi.org/10.1080/08878730.2018.1516260

Mehan, H. (1998). The study of social interaction in educational settings: Accomplishments and unresolved issues. *Human Development, 41*, 245-269.

Meiers, M., Ozolins, C., & Mckenzie, P. (2007). Improving consistency in teacher judgements : an investigation for the Department of Education, Victoria.

Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment in kindergarten to grade 3. *American Educational Research Journal, 38*(1), 73-95. https://doi.org/10.3102/00028312038001073

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (pp. 13-103). American Council on Education and Macmillan

Messick, S. (1998). Test Validity: A Matter of Consequence. *Social Indicators Research, 45*(1-3), 35-44. https://doi.org/10.1023/a:1006964925094

Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice, 22*(4), 13-25. https://doi.org/10.1111/j.1745-3992.2003.tb00140.x

Murillo, F. J., & Hidalgo, N. (2017). Students' conceptions about a fair assessment of their learning. *Studies in Educational Evaluation, 53*, 10-16. https://doi.org/https://doi.org/10.1016/j.stueduc.2017.01.001

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis:Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods, 16*(1), 1609406917733847. https://doi.org/10.1177/1609406917733847

Palmer, S. (2004). Authenticity in assessment: Reflecting undergraduate study and professional practice. *European Journal of Engineering Education, 29*(2), 193-202. https://doi.org/10.1080/03043790310001633179

Pepper, M. B., & Pathak, S. (2008). Classroom contribution: What do students perceive as fair assessment? *Journal of Education for Business, 83*(6), 360-367. https://doi.org/10.3200/JOEB.83.6.360-368

Rasooli, A., Zandi, H., & DeLuca, C. (2018). Re-conceptualizing classroom assessment fairness: A systematic meta-ethnography of assessment literature and beyond. *Studies in Educational Evaluation, 56*, 164-181. https://doi.org/https://doi.org/10.1016/j.stueduc.2017.12.008

Scott, S., Webber, C. F., Lupart, J. L., Aitken, N., & Scott, D. E. (2014). Fair and equitable assessment practices for all students. *Assessment in Education: Principles, Policy & Practice, 21*(1), 52-70. https://doi.org/10.1080/0969594X.2013.776943

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagnè, & M. Scriven (Eds.), *Perspectives of Curriculum Evaluation*. Rand McNally.

Shepard, L. A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of Research in Teaching* (4th ed., pp. 1066-1101). American Educational Research Association.

Siegel, M. A. (2007). Striving for equitable classroom assessments for linguistic minorities: Strategies for and effects of revising life science items. *Journal of Research in Science Teaching, 44*(6), 864-881. https://doi.org/https://doi.org/10.1002/tea.20176

Smith, J. K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice, 22*(4), 26-33. https://doi.org/10.1111/j.1745-3992.2003.tb00141.x

Snekalatha, S., Marzuk, S. M., Meshram, S. A., Maheswari, K. U., Sugapriya, G., & Sivasharan, K. (2021). Medical students' perception of the reliability, usefulness and feasibility of unproctored online formative assessment tests. *Advances in Physiology Education, 45*(1), 84-88. doi:10.1152/advan.00178.2020

Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education: Principles, Policy & Practice, 12*(3), 275-287. https://doi.org/10.1080/09695940500337249

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Pearson Education Inc.

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing, 30*(3), 403-412. https://doi.org/10.1177/0265532213480338

Thorndike, E. L. (1918). The nature, purposes, and general methods of measurement of educational products. In S. A. Courtis (Ed.), *The Measurement of Educational Products (17th Yearbook of the National Society for the Study of Education, Pt. 2)* (pp. 16-24). Public School

Tight, M. (2012). Higher education research 2000–2010: changing journal publication patterns. *Higher Education Research & Development, 31*(5), 723-740. https://doi.org/10.1080/07294360.2012.692361

van der Vleuten CP, Schuwirth LW. (2005). Assessing professional competence: From methods to programmes. Med Educ 39:309–317.

van der Vleuten C.P, Schuwirth L.W., Driessen, E.W. Govaerts, M.J.B. & Heeneman, S. (2015). Twelve tips for programmatic assessment. *Medical Teacher, 37*, 641-646.

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education, 43*, 840 - 854.

Webb, C., Endacott, R., A Gray, M., Jasper, M. A., McMullan, M., & Scholes, J. (2003). Evaluating portfolio assessment systems: what are the appropriate criteria? *Nurse Education Today, 23*(8), 600-609. https://doi.org/https://doi.org/10.1016/S0260-6917(03)00098-4

Wiliam, D. (1993). Validity, dependability and reliability in National Curriculum assessment. *The Curriculum Journal, 4*(3), 335-350. https://doi.org/10.1080/0958517930040303

Yung, B. H. W., & Yung, B. H. W. (2001). Three views of fairness in a school-based assessment scheme of practical work in biology. *International Journal of Science Education, 23*(10), 985-1005. https://doi.org/10.1080/09500690010017129