

The validity, reliability, academic integrity and integration of oral assessments in higher education: A systematic review

Shashi Nallaya, Sheridan Gentili, Scott Weeks and Katherine Baldock

University of South Australia, Adelaide, Australia

Various factors such as regulatory body mandates, graduate employability challenges, decreasing student engagement and increasing academic misconduct in higher education have motivated universities to explore alternative approaches to teach and assess. Accordingly, the oral assessment has taken precedence in many contexts as a popular form of assessment. Although literature highlights the strengths of oral assessments, there are also reservations among some scholars as it is found to contribute to issues pertaining to validity and reliability. This paper reports on a systematic review undertaken in accordance with the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) guidelines to explore the extent of oral assessment validity, reliability and capacity to address academic misconduct in higher education. A total of 2,657 journal articles from ERIC, Web of Science, Scopus and A+ Education databases were imported into *Covidence* for screening of titles, abstracts and full texts. Seventeen studies were deemed suitable for inclusion in this systematic review. The analysis identified that the validity, reliability and capacity of the oral assessment to reduce academic integrity breaches were dependent on whether it has been designed, scaffolded, and implemented well.

Introduction

Factors such as regulatory body recommendations (Tertiary Education Quality and Standards Agency, 2015), graduate employability challenges (James & Casidy, 2018; Sotiriadou et al., 2020), decreasing student engagement (Hart et al., 2011) and increasing academic integrity concerns in higher education (Thomas, Raynor & McKinnon, 2014) have motivated universities to explore alternative approaches to teach and assess. Research in this space has identified that authentic assessments have the capacity to equip students with the skills and knowledge required in professional settings, increase graduate employability, improve student engagement in learning and teaching activities as well as reduce academic integrity breaches (Ashford-Rowe et al., 2014; Villarroel et al., 2018). Authentic assessments are those assessments that align classroom learning with the workplace, mimicking tasks and performance standards that would be expected of professionals (Villarroel et al., 2018; Wiggins, 1990).

In exploring more creative and authentic ways of assessing students, and improving assessment security, the oral assessment has taken precedence as a form of authentic assessment used both in formative and summative contexts (Akimov & Malin, 2020; Sutherland et al., 2019). In a quick review of literature on oral assessment, the authors identified that oral assessments were used in various settings both nationally and internationally (for example, Giordano & Christopher, 2020; Iannone & Simpson, 2015; Sutherland et al., 2019). Although most of these studies reported an improvement in student performance in the assessment, other studies also highlighted some limitations of

oral assessment (see Bhati, 2012; Turner et al., 2014). Lack of standardisation, equity problems for students speaking English as an additional language (EAL) and student anxiety were some limitations reported in the findings of these studies. The initial review of literature contributed towards the curation of two research questions: (1) What is the validity and reliability of oral assessments as an alternative form of authentic assessment in higher education? and (2) What is the role of oral assessments in addressing academic integrity challenges in higher education? Our article reports on a systematic review undertaken to find answers to the two research questions. Key themes relevant to the systematic review are discussed with reference to literature to provide context to the questions under investigation.

Authentic assessment

Assessments play a central role in any educational environment as they influence how students will engage in the content and activities. However, there is much concern in educational settings around the world about the validity and reliability of assessments that solely test students' ability to recall knowledge, as demonstrated by some graduates who appear to be high achievers on their exam scripts but fail in their ability to perform in the workplace. Standardised assessments, it seems, do not provide opportunities to assess students' true ability to undertake and solve real world challenges (Pereira et al., 2016). Wiggins (2011, p. 84), one of the advocates of authentic assessments, argued that the "true test of ability is to perform consistently well [in] tasks whose criteria for success are known and valued". Assessments that demonstrate this are deemed authentic assessments, i.e., tasks that resemble problems that exist or emulate those in a discipline or the real world. The problems identified in authentic tasks are complex and messy, requiring students to think from multiple perspectives and use higher order thinking to develop solutions (Ashford-Rowe et al., 2014; Villarroel et al., 2018; Wiggins, 2011).

It is pertinent that study programs equip students with the skills and knowledge that will make them job ready as increasingly university graduates are unable to find work in their area of study within a year of graduating (Sotiriadou et al., 2020). The need for authentic assessment tasks is also becoming progressively important, specifically in the current higher education environment where students study through multiple modalities and geographic locations (Akimov & Malin, 2020), resulting in a growing concern about assessment security and the increase in academic misconduct and contract cheating (Dawson, 2020). However, it is worth knowing that authentic design of assessments alone will not prevent the prevalence of academic misconduct as highlighted by Ellis et al. (2020). Nevertheless, thoughtful design can make cheating more difficult. When designing assessments, more thought needs to be invested so that students perceive them to be authentic, meaningful and of value. Students' language capabilities and academic literacies should also be developed and scaffolded so that they are confident and capable enough to undertake and engage in their assessment tasks. Students must also be instructed about academic integrity from the onset of their study program to know what behaviours are unacceptable when completing their assessments.

Oral assessment as a form of authentic assessment

Students will engage in assessment tasks which they perceive as necessary and of value. In addition, and in response to increasing calls to address workplace graduate employability, institutions are putting in measures to ensure that students graduate with the skills and knowledge required by employers (James & Casidy, 2018; Sotiriadou et al., 2020). Now, more than in the past, there is a need for assessments and teaching and learning activities to reflect elements of authenticity so that students are prepared for the workplace and equipped with the necessary knowledge and skills.

The oral assessment's (sometimes referred to as *viva voce* or *mini viva*) popularity and use has increased in more recent times in undergraduate studies and certification-type examinations, after gaining recognition for its ability to assess students' deep understanding and knowledge, critical thinking and reduce the likelihood of academic misconduct (Salamonson et al., 2016). The oral assessment also has the potential to examine a student's workplace readiness and professionalism. Furthermore, oral assessments are favoured over other, more traditional forms of assessments in many contexts (see studies analysed in the systematic review). This assessment type allows examiners to interact with students to identify their strengths and distinguish between superficial knowledge and deep understanding (Pearce & Lee, 2009). It is reported that oral communication skills are viewed as the most important or at least one of the skills prospective recruiters or employers look for in a potential employee (Brink & Costigan, 2015), hence making the oral assessment an authentic form of assessment.

The oral assessment is defined as an "assessment in which a student's response to the assessment task is verbal, in the sense of being expressed or conveyed by speech instead of writing" (Joughin, 1998, p. 367). It is also defined as a situation in which the candidate gives spoken responses to questions from one or more examiners (Huxham, Campbell & Westwood, 2012). The oral assessment is used as an authentic assessment in many disciplines, such as medicine, nursing, marketing and finance as this assessment can provide conditions for emulating as closely as possible, situations or contexts where two-way interaction is required in the provision of service (see for example Pearce, 2009; Salamonson et al., 2016; Sutherland, 2019), thus mimicking students' future workplace.

Although literature suggests that many students found the oral assessment to be an assessment type they favoured (Koh et al., 2021; Thomas et al., 2014), it was also highlighted that it made some students anxious. The reliability of such an assessment also came into question as some students were worried that the examiners and question types might be biased. Additionally, it was emphasised that since students in some educational settings had not experienced or trained for the oral assessment, they did not know what to expect, thus contributing to heightened anxiety (Burke-Smalley, 2015; Iannone & Simpson, 2015; Thomas et al., 2014). Although some of the reviewed studies referenced anxiety as a limitation of the oral assessment, the authors had also indicated that students' scores in their oral assessment was higher than their written assessment (Burke-Smalley,

2015), students perceived it to be relevant to their future career (Iannone & Simpson, 2015) and that students preferred it to written exams (Thomas et al., 2014).

Methods

This systematic review is reported in accordance with the criteria set out in the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) (Page et al., 2021).

Search strategy

The electronic databases ERIC, Web of Science, Scopus and A+ Education were systematically searched to identify relevant published peer-reviewed studies. An initial literature review was conducted to develop keywords for the concept (oral assessment) and then further refined using index terms from the electronic databases searched to develop the full search strategy. Search terms for population (students undertaking oral assessments) and context (university undergraduate studies) were not added to the search strategy to ensure potentially relevant articles were not missed. Figure 1 presents the search string conducted in ERIC. Keywords and index terms in the search strategy were then adapted to allow for syntax variations between databases. The search was limited to studies published in English and between 2010-2021 to keep the findings relevant and applicable to the current higher education context. The first 200 citations in Google Scholar (“oral assessment” OR “oral examination” OR “oral test” OR “oral quiz” OR “oral presentation” or “oral online” OR “viva Voce” OR eViva OR “mini viva” OR digital viva”) between 2010-2021, as well as the reference lists of all included studies were screened by title to ensure potentially relevant publications were not missed in the database search. The systematic search was validated by an academic librarian.

(Oral NEAR/2 (assessment OR exam* OR test OR quiz OR presentation OR online)) OR (Viva NEAR/2 (voce OR exam* OR assessment OR online)) OR eViva OR (Mini) AND (viva) OR (Digital) AND (viva)

Figure 1: ERIC search string

Eligibility criteria for included studies

Studies reporting on any type of oral assessment in undergraduate higher education were included in this review. Other types of assessment (e.g., written assignments, exams, portfolios, reports) at an undergraduate level, or oral assessments in postgraduate higher education, vocational education and training (VET) and secondary school were excluded. This review included peer-reviewed publications with quantitative and qualitative study designs. Literature reviews, dissertations, conference abstracts, books, teaching and learning guides and white papers were excluded.

Study selection

All identified records from the electronic database search were collated and uploaded into *EndNote20* (Clarivate Analytics, USA (<https://www.endnote.com>) and then imported to *Covidence* (Veritas Health innovation, Australia, <https://www.covidence.org>) for screening. Duplicate records were removed in Covidence. Following a pilot test of the first 100 articles to revise the inclusion criteria, all articles were screened by title and abstract by two independent reviewers. The full texts of potentially relevant articles were retrieved and assessed against the inclusion criteria by two independent reviewers. Reasons for exclusion were reported for full-text studies. Disagreements between reviewers at each stage of study selection were resolved through discussion to reach consensus, or with a third reviewer. The results of the screening process are reported in a *Preferred Reporting Items for Systematic Reviews and Meta-analyses* (PRISMA) flow diagram (Page et al., 2021).

Data extraction and synthesis

Data were extracted from included studies by two independent reviewers in Covidence. Data were extracted on author, date, country, professional discipline, and data relating to the type, validity, reliability, academic integrity, and outcomes of “oral assessment”. Discrepancies were resolved through discussion to reach consensus, or with a third reviewer. Data are presented as a narrative summary and in a table of study characteristics (Table 1).

Results

A total of 2657 records were identified through the electronic database search. After duplicates were removed, 2199 records were screened by title and abstract, of which 368 received full text review. There were 17 studies that met the inclusion criteria and are included in this review. All included studies were peer-reviewed and written in English. No additional studies were found on Google Scholar or in reference lists of included studies. Figure 2 presents the screening results and reasons for exclusion at full text review in a PRISMA flow diagram.

Seventeen (n=17) studies were deemed suitable for inclusion in the systematic review (Badger, 2010; Burke-Smalley, 2014; Dicks et al., 2012; Hazen, 2020; Huxham et al., 2012; Iannone et al., 2020; Iannone, 2012; Kang et al., 2019; Koh et al., 2021; Luckie et al., 2013; Salamonson et al., 2016; Simper, 2010; Sotriadou et al., 2020; Thomas et al., 2014; Turjamaa et al., 2018; Turner & Davila-Ross, 2015; Wallace, 2010). Five of these studies were undertaken in the United States, six in the United Kingdom, two in Australia, and one each in Singapore, Canada, United Arab Emirates and Finland, suggesting that attempts to administer authentic forms of assessment are a common practice around the world.

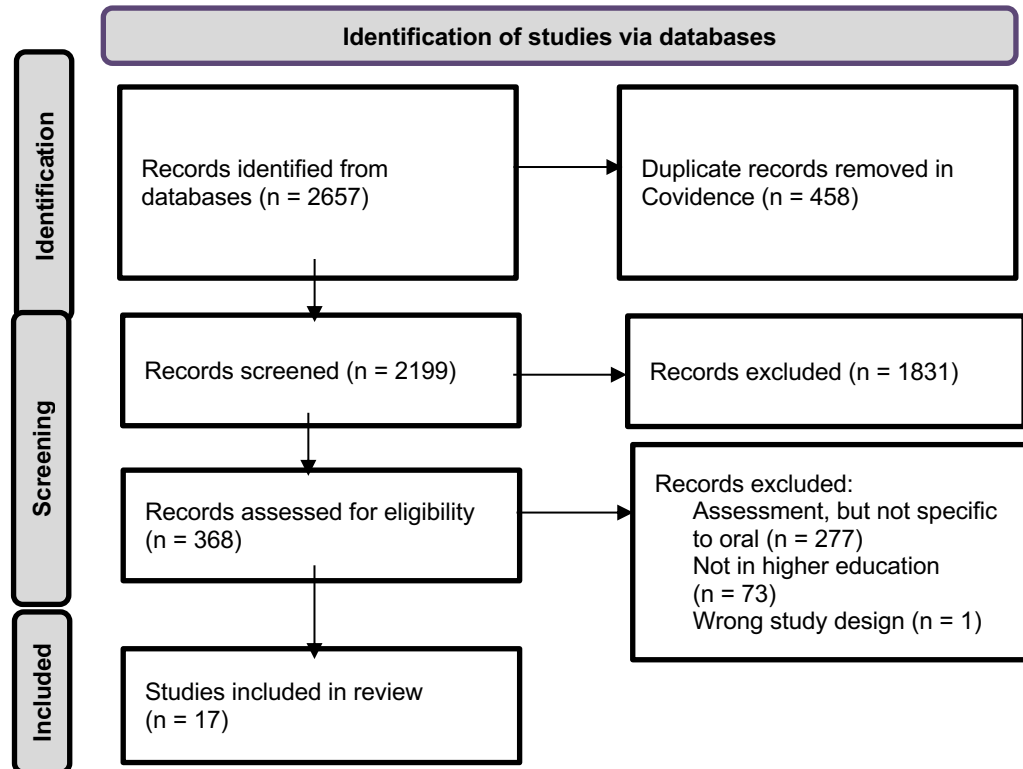


Figure 2: Prisma flow diagram

Table 1 summarises the studies analysed in the systematic review. Publication dates ranged from 2010 to 2021 identifying that oral assessments have been topical and popular over the last decade. Despite the abundance of literature on oral assessment, the interest among educators to explore more authentic approaches to administer the assessment does not seem to wane. There were a range of disciplines where the oral assessments were implemented including education, business, chemistry, geography, biology, mathematics, diversity and race, nursing, advanced exercise nutrition, psychology and law, reflecting the versatility of the assessment.

Although the fundamental principles of the assessment were the same, the reviewed studies highlighted that they were referred by various names, the most common being oral exam or oral examination (n=4). To answer the two research questions, the studies in this review were analysed particularly for any references made to validity, reliability and academic integrity considerations besides identifying the outcome of the interventions.

Table 1: Summary of the studies analysed in this systematic review

Study	Country	Discipline	Name	Validity	Reliability	Academic integrity	Outcome
Badger (2010)	USA	Education	Summative oral exam	Yes			Yes
Burke-Smalley (2014)	USA	Business	Oral quiz		Yes	Yes	Yes
Dicks et al. (2012)	Canada	Organic Chemistry	Oral exam		Yes	Yes	Yes
Hazen (2020)	USA	Geography	Oral exam		Yes	Yes	Yes
Huxham et al. (2012)	UK	Biology field methods	Oral exam	Yes			Yes
Iannone et al. (2020)	UK	Maths	Oral performance assess.	Yes	Yes		Yes
Iannone (2012)	UK	Maths	One to one tutorial			Yes	Yes
Kang et al. (2019)	USA	Diversity and race	Oral exam	Yes	Yes	Yes	Yes
Koh et al. (2021)	Sing	Chemistry	Viva voce				Yes
Luckie et al. (2013)	USA	Biology	Verbal final exam		Yes		Yes
Salamonson et al. (2016)	Aust	Nursing	Oral viva assessment	Yes			Yes
Simper (2010)	UK	Advanced exercise nutrition	Oral viva exam		Yes	Yes	Yes
Sotiriadou et al. (2020)	Aus	Business	Interactive oral assess.			Yes	Yes
Thomas et al. (2014)	UAE	Psychology	Group-based oral exam	Yes		Yes	Yes
Turjamaa et al. (2018)	Finland	Nursing	Dialogic group oral exams	Yes	Yes		Yes
Turner & Davila-Ross. (2015)	UK	Psychology	Oral project interviews	Yes	Yes	Yes	Yes
Wallace (2010)	UK	Law	One to one oral assess.	Yes			Yes

Note: "Yes" refers to references made to validity, reliability, academic integrity and intervention outcomes

Validity of oral assessments

Validity can be defined as the degree to which “an assessment measures what it is essentially designed to measure” (Knapp & Mueller, 2010, 337). Ten (n=10) of the studies reviewed, referred to the validity of oral assessments (Badger, 2010; Hazen, 2020; Huxham et al., 2012; Iannone et al., 2020; Kang et al., 2019; Salamonson et al., 2016;

Thomas et al., 2014; Turjamaa et al., 2018; Turner & Davila-Ross, 2015; Wallace, 2010). Badger (2010) identified that the oral assessment is not authentic in that it lacks many factors that would be present in real contexts. In addition, some studies identified that students' performance in the assessment might be compromised because of the anxiety that this form of assessment induces (Hazen, 2020; Huxam et al., 2012; Iannone et al., 2020; Kang et al., 2019; Thomas et al., 2014).

Turner and Davila-Ross's study (2015), however, reported no significant differences between students' performance in their oral assessment and written project report, suggesting that although the oral assessment may induce stress and anxiety in some students, it does not significantly impact on their performance when compared to more traditional written assessments (Turner & Davila-Ross, 2015). The potential negative impact of oral assessments on students from EAL backgrounds was highlighted as one of the main disadvantages of this form of assessment (Hazen, 2020; Kang et al., 2019; Salamonson et al., 2016). Conversely, most of the respondents (89%) in Hazen's (2020) study indicated that their spoken English proficiency had not been an important consideration for them in undertaking the oral assessment. The reviewed studies also suggested that the oral assessment would be more suitable for some personalities compared to others (extroverts v introverts). For instance, students who were shy or not confident speaking in public would not perform well in this type of assessment (Huxham et al., 2012; Turjamaa et al., 2018; Turner & Davila-Ross, 2015). Only one of the studies suggested that the oral assessment may not be fair and that respondents may not be clear about the assessed criteria or have the necessary experience or skills to adequately prepare for the assessment. (Wallace, 2010).

Reliability of oral assessments

Reliability can be defined as the consistency of the measurement from "time to time, form to form, item to item or one rater to another" (Knapp & Mueller, 2010, p. 337). Nine (n=9) out of the 17 screened studies had referred to the reliability of oral assessments (Burke-Smalley, 2014; Dicks et al., 2012; Hazen, 2020; Iannone et al., 2020; Kang et al., 2019; Luckie et al., 2013; Simper, 2010; Turjamaa et al., 2018; Turner & Davila-Ross, 2015). Many of these studies had acknowledged concerns about the oral assessment's reliability by identifying measures in the assessment design to improve on this aspect. The measures taken to enhance reliability of the oral assessment included predetermining the allocated time, asking each student the same question, and establishing assessment criteria prior to the assessment (Burke-Smalley, 2014). Reliability concerns were also reduced through the use of rubrics (Burke-Smalley, 2014; Kang et al., 2019).

Interrater reliability was another concern associated with oral assessments (Burke-Smalley, 2014; Dicks et al., 2012; Hazen, 2020; Turner & Davila-Ross, 2015). Interrater reliability of oral assessments in the reviewed studies was addressed by having more than one assessor present (Dicks et al., 2012). However, Turner and Davila-Ross (2015) have demonstrated that having two or more assessors does not necessarily overcome challenges associated with interrater reliability. Their study found significant differences between the marks awarded by the first and second assessor. Additionally, a minority group of students

(n=7) in Hazen's (2020) study indicated that they perceived the oral exam to be unfair and six per cent of students in the same study suggested that the grading of the exam would be subjective. These students' views add to the reliability concerns of the oral assessment. Similarly, Iannone et al. (2020) also questioned the reliability of assessments that focused solely on students' written output as this would limit the demonstration of their actual capability.

Luckie et al. (2013) suggested that the oral assessment was reliable as they did not find any significant differences related to student satisfaction of the assessment regarding age or gender. These researchers also pointed out that generally, students who scored higher grades in the oral assessment were more satisfied compared to those with lower grades. They also found that EAL students were generally less confident with oral assessments, strengthening the validity concerns reported above, which suggested that this assessment type may disadvantage EAL students.

Furthermore, Simper's (2010) study draws attention to another reliability concern associated with oral assessment. Respondents who participated in the study suggested that student success in the oral assessment could be dependent on their skills and personality. This affirmed the earlier findings related to validity, which highlighted that confident and extroverted students will perform better than shy and introverted students.

Oral assessment and academic integrity

Eight (n=8) out of the seventeen studies analysed had referred to academic integrity (Burke-Smalley, 2014; Dicks et al., 2012; Iannone, 2012; Kang et al., 2019; Simper, 2010; Sotiriadou et al., 2020; Thomas et al., 2014; Turner & Davila-Ross, 2015). Some of the studies suggest that oral assessments can mitigate academic misconduct (Burke-Smalley, 2014; Dicks et al., 2012; Iannone, 2012; Kang et al., 2019; Sotiriadou et al., 2020; Thomas et al., 2014; Turner & Davila-Ross, 2015). The analysis identified that in some instances, the oral assessment was administered to address academic misconduct cases reported in other forms of assessment previously offered (Kang et al., 2019; Simper, 2010).

From the analysis, it can also be inferred that the oral assessment format allows students to work together while simultaneously staying clear of behaviours such as collusion, which can result in breach of institutional academic integrity policies and procedures (Dicks et al., 2012; Thomas et al., 2014). Furthermore, the student respondents in the analysed studies highlighted that it would be difficult to engage in contract cheating services in some oral assessments such as the interactive oral assessment and oral project interviews (Sotiriadou et al., 2020; Thomas et al., 2014; Turner & Davila-Ross, 2015), suggesting that assessment design can reduce the likelihood of academic misconduct in summative assessment.

Outcomes of oral assessment interventions

The analysis of the studies identified that, to an extent students were anxious about oral assessments as they were not familiar with this type of assessment. In addition, students

had previously experienced only paper-based tests (Badger, 2010; Hazen, 2020; Koh et al., 2021; Thomas et al., 2014). Interestingly, the analysis suggests that most students performed better in oral assessments than they did in written assessments (Burke-Smalley, 2014; Dicks et al., 2012; Huxham et al., 2012; Luckie et al., 2013) and that this assessment inspired them to engage in more thorough preparation (Badger, 2010; Dicks et al., 2012; Kang et al., 2019; Turjamaa et al., 2018). Conversely, some students demonstrated similar performance in both their oral and written assessment (Hazen, 2020).

It was found that oral assessments provided students with the opportunity to demonstrate a deeper understanding of the course content (Badger, 2010; Iannone, 2012; Luckie et al., 2013; Turjamaa, 2018; Wallace, 2010), as there was a likelihood for students to forget information they had acquired in closed-book assessments that required students to memorise information (Iannone et al., 2020). Some of the respondents in the analysed studies suggested that the oral assessment was an authentic form of assessment as it was more relevant to the requirements of the workplace (Iannone et al., 2020; Sotiriadou et al., 2020; Thomas et al., 2014; Turjamaa et al., 2018). The findings also suggested that the oral assessment enhanced the final grades of weaker students (Dicks et al., 2012) and that students who completed the oral assessment generally achieved better grades overall (Burke-Smalley, 2014; Dicks et al., 2012; Luckie et al., 2013; Turner & Davila-Ross, 2015).

Despite the many positive attributes identified in the review of the studies, the analysis highlighted that the oral assessment does not meet the learning needs of some students, demonstrating better performance in their written assessment than their oral assessment (Hazen, 2020). Anxiety appeared to be the most significant variable that was reported in the analysis (Iannone et al., 2020; Kang et al., 2019; Thomas et al., 2014), affirming the findings of the earlier literature review. Interestingly, in Koh et al.'s (2021) study, the respondents indicated feeling less anxious when sitting for the oral assessment in the 'viva la vida' approach. The respondents were referring to the design of the oral assessment. In this study, the researchers conducted a gamified viva voce in their third-year chemistry laboratory to foster communication and teamwork through collaborative learning (Koh et al., 2021, p. 2018). This finding suggests that the design of oral assessments needs careful consideration for students to positively engage in the assessment. Accordingly, Simper (2010) found no evidence in their research to clearly identify the impact of anxiety on the oral assessment, compared to previous years' paper-based assessment.

The analysis also highlighted that some students, for example those from EAL backgrounds, may be disadvantaged by this type of assessment (Salamonson et al., 2016). With reference to other variables such as gender, it was highlighted that there was no statistical significance between male and female performance in the oral assessment, or student age. There was, however, a difference between overall performance in oral assessment compared to written assessments independent of gender (Burke-Smalley, 2014; Huxham et al., 2012). Salamonson et al.'s (2016) study found no differences in student satisfaction of the oral assessment independent of gender and age.

Analysis showed no clear demarcation with regard to student preference for oral assessments and written assessments, with some students preferring the former while

others preferred the latter (Iannone et al., 2020; Kang et al., 2019; Koh et al., 2021; Salamonson et al., 2016; Simper, 2010; Thomas et al., 2014; Wallace, 2010). Analysis of the studies also highlighted the need for careful consideration when designing oral assessments to ensure positive outcomes for students. For example, the assessors need to be skilful in ensuring that the two-way interaction progresses efficiently for both proficient and weak students, within the allocated time (Thomas et al., 2014). It was also important for the assessors to create a safe and supportive environment so that students were motivated to engage in the discussion (Turjamaa et al., 2018). In Wallace's (2010) study, it was found that students could not clearly identify how they prepared for the oral assessment, however, everyone agreed that it would be easier than the written assessment.

Discussion

The systematic review discussed in this paper aimed to find answers to two research questions, namely (1) What is the validity and reliability of oral assessments as an alternative form of authentic assessment in higher education? and (2) What is the role of oral assessments in addressing academic integrity challenges in higher education? The analysis identified that the oral assessment's validity, reliability and capacity to reduce academic integrity breaches would depend on its design (Akimov & Malin 2020). When designing an assessment, educators need to firstly identify the learning objectives they want to assess and whether the chosen assessment type is the best method to assess the learning objectives they have chosen to assess, thus contributing to the validity of the assessment. For example, if assessors want to test students' in-depth knowledge, problem-solving, decision-making, communication and ethical reasoning skills; qualities that would be expected of a graduate in professional settings, then the oral assessment can play a significant role as an assessment type, especially in some disciplines where graduates are expected to interact with their colleagues, clients and society in their day-to-day dealings (Badger, 2010; Dicks et al., 2012; Thomas et al., 2014). Assessors also need to ensure that the topics being tested in the assessment have been covered in class and a descriptive and detailed rubric is used during the assessment to ensure objectivity during the assessment. The rubrics should also be presented to the students prior to the assessment so that they know what aspects will be assessed during the assessment and can prepare appropriately. If these aspects are clearly considered and incorporated in the design, then the oral assessment can be a valid form of authentic assessment (Iannone & Simpson 2014).

Due to the subjective nature of oral assessments, the analysis of the articles appraised in this systematic review identified concerns regarding its reliability. To enhance reliability of oral assessments, assessors need to undergo training to ensure that they are objectively assessing the students and consistently awarding similar marks. Although increasing the number of assessors does not necessarily contribute to inter-rater reliability (Davila-Ross, 2015), effective training and moderation during and post-assessment can reduce concerns regarding subjectivity of the oral assessment (Dicks et al., 2012). Additionally, consistent opportunities for students to practise in class for the assessment, and availability of good resources that will equip them with skills to sit for the oral assessment, can also contribute to its reliability. As with validity, students' concerns with the reliability of the oral

assessment can be reduced if they are adequately prepared and not surprised by any element that they had not envisioned during the assessment. A clear, descriptive and detailed rubric that sets out expectations can reduce issues pertaining to reliability (Burke-Smalley, 2014; Kang et al., 2019). Likewise, the oral assessment should not focus only on students' written output (Iannone et al., 2020) but should also provide opportunities for demonstration of their overall capabilities, mimicking future workplace settings.

The systematic review also highlighted that the oral assessment does not suit all students, particularly those who are shy or who are not comfortable speaking in public. It was also identified that this type of assessment can disadvantage students from EAL backgrounds (Huxham et al., 2012; Turjamaa et al., 2018; Turner & Davila-Ross, 2015). For these students, more support and scaffolding should be provided so that they are comfortable and confident to sit for the oral assessment. Communication skills are one of the top capabilities that employers look for in graduates. In this aspect, the oral assessment is an authentic assessment type (Thomas et al., 2014; Turjamaa et al., 2018). Shying away from the oral assessment because it disadvantages some students would result in students not being job ready. Conversely, teachers need to instil in their students the relevancy of communication skills for future workplace and develop their skills to master this graduate attribute.

To an extent, the oral assessment can mitigate academic misconduct in higher education (Kang et al., 2019; Sotiriadou et al., 2020), as it has this capacity, often missing in other forms of standardised or paper-based assessments that motivate students to engage in academic misconduct. If designed well, for example in the interactive oral assessment and oral project interviews, it would be difficult for students to engage in contract cheating (Thomas et al., 2014; Turner & Davila-Ross, 2015). As far as possible, there should be a face-to-face element involved in the oral assessment. If this is not possible due to class size and geographic location of students, then assessors need to ensure that appropriate protocols and proctoring are implemented so that students cannot engage in academic misconduct, hence contributing to the assessment's academic integrity.

In summary, the systematic review discussed in this paper highlights some important factors for consideration. These include integrated and holistic approaches to assessment while addressing the limitations of a 'one size fits all' approach to assessing students. Aspects such as learning objectives and the purpose for assessing students, should always direct assessment design, to ensure a comprehensive evaluation process. Educators should recognise that a single assessment type may not be suitable for all students. When selecting assessment type, there should be consideration for diversity of learning needs and learning styles. The authors of this paper advocate for a thoughtful and inclusive approach to assessment, while recognising the potential of well-designed oral assessment in addressing academic integrity and fostering essential skills for future success in the workplace.

Implications for higher education

Recently, globally, the higher education sector has been rocked by the emergence of generative artificial intelligence (GenAI) technology. For many this was a wakeup call, whilst for others there was recognition of the emergence of artificial intelligence in learning and what it would mean to the higher education sector. The speed at which GenAI tools have become ubiquitously adopted across all part of society has surprised many. Academics are increasingly questioning the rigour of assessment design, and in the post-Covid education environment, which saw much of our face-to-face teaching and learning, and assessment delivery, move online, querying the authenticity of assessments submissions. This becomes critical for discussion as the sector uses assessment as a means to demonstrate attainment of learning competences and graduate professional standards. The implications of such disruptive technologies will (and is already) leading to universities possibly conferring degrees to students who have not demonstrated the learning outcomes and therefore who may not have acquired the skills and knowledge necessary for their workplace.

Oral assessments provide an opportunity, when rigorously designed and implemented, to address the growing concerns raised by current and potentially future disruptions to the higher education sector. Furthermore, they provide opportunity to address the inherent bias of our education model and assessment practices by enabling all students, irrespective of their ability, to engage with their assessors with appropriate support during the assessment task through prompting questions. When designed and delivered well, oral assessment can create a safe space for all students, reducing some of the access and equity barriers which confront students and impact their learning.

Traditional assessments such as essays, oral presentations, or quizzes, are used, as discussed previously, to validate our assumptions as to whether discipline knowledge has been attained. Fostering the development and assessment of human skills such as emotional intelligence, creative problem solving, critical thinking, interpersonal skills, cultural competencies, ethical decision making, adaptability, leadership, empathy, and communication can be more challenging. Forms of assessments that clearly evaluate a student's ability to critically think and analyse information in real time further strengthens the sector's role in educating future professionals. The systematic review of literature discussed in this paper has suggested that through authentic assessment design, oral assessment can provide genuine opportunities to support students develop core human skills, which will be increasingly required moving forward, hence meeting the needs and expectation of professional practice and society more broadly.

Holistic implementation of new approaches to assessment design and delivery does have challenges. Research informed assessment design which acts to support industry and societies' expectations regarding the development of communication and other human skills, reinforces the central role that universities play in educating our future professionals. Whilst workload concerns are often cited as factors limiting the development of novel forms of assessment, including the development and delivery of

oral assessments, the overall positive impact on student engagement, learning, and academic integrity, should outweigh these concerns.

Conclusion

This paper presents what the authors believe to be the first systematic review to bring together the findings from across the higher education sector to present a definitive position on the value of oral assessments, student views, and considerations that can form the basis of a framework for improved integration of oral assessment. The authors also advocate for more diverse assessment types so that by the time students graduate from a study program, they have acquired all the knowledge, skills, competences, qualities and values in preparation for further studies or lifelong learning.

The studies analysed here, which were systematically selected based on defined selection criteria, present several considerations for the successful development and implementation of oral assessment within a course. None of the articles presented, discussed systematic implementation across a whole degree or institution. This is important as oral assessments, as a form of formative and summative assessment, should be scaffolded through the student's learning journey. This may occur entirely in a course, or across the entire program of learning. Designed and implemented this way, the oral assessment has the potential to provide ample opportunity for students to develop the necessary skills, whilst simultaneously engaging in discipline learning. What this form of design and delivery necessitates is deliberate space be created within the curriculum, in addition to spaces used for practice (physical or virtual including software). To guide student learning, clear expectations must be provided to students via the assessment rubrics, something we consider to be gold standard in all assessment design. Early and clear communication along with regular opportunities for practice regarding assessment expectations act to minimise the development of anxiety over the teaching period.

Critical to oral assessments is academic staff professional development – one could argue that this is more important for oral assessment than other forms of assessment. When delivered well, oral assessment provides students with an opportunity to demonstrate their learning through structured and unstructured open-ended questions. Whilst the delivery of structured questions can be straight forward, the delivery of unstructured questions may be more challenging, as they must not inadvertently lead students toward the desired answer. To facilitate the development and delivery of oral assessment questions, academics should engage with industry professionals and/or industry advisory groups. Lastly, robust moderation and review practices (both institutional and developed as part of assessment delivery) are important to ensure that the assessment delivery is valid, fair and reliable.

The authors conducted the systematic review to understand key considerations for the implementation of oral assessment from existing literature. The delivery of authentic forms of learning and assessment have become increasingly popular over recent years, driven primarily by the shift into online learning and the growing concerns around

academic integrity. The context of how we work as a society has shifted significantly over the last five years and will continue to evolve in unknown ways over the next five to ten years. This necessitates a deliberate and considered shift towards forms of assessment to support the development of future skills, whilst staying true to our core values as higher education institutions.

References

- Akimov, A. & Malin, M. (2020). When old becomes new: A case study of oral examination as an online assessment tool. *Assessment & Evaluation in Higher Education*, 45(8), 1205-1221. <https://doi.org/10.1080/02602938.2020.1730301>
- Ashford-Rowe, K., Herrington, J. & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205-222. <https://doi.org/10.1080/02602938.2013.819566>
- Badger, J. (2010). Assessing reflective thinking: Pre-service teachers' and professors' perceptions of an oral examination. *Assessment in Education: Principles, Policy & Practice*, 17(1), 77-89. <https://doi.org/10.1080/09695940903076022>
- Bhati, S. S. (2012). The effectiveness of oral presentation assessment in a finance subject: An empirical examination. *Journal of University Teaching & Learning Practice*, 9(2), article 6. <https://doi.org/10.53761/1.9.2.6>
- Brink, K. E. & Costigan, R. D. (2015). Oral communication skills: Are the priorities of the workplace and AACSB-accredited business programs aligned? *Academy of Management Learning & Education*, 14(2), 205-221. <https://www.jstor.org/stable/43697262>
- Burke-Smalley, L. A. (2014). Using oral exams to assess communication skills in business courses. *Business and Professional Communication Quarterly*, 77(3), 266-280. <https://doi.org/10.1177/2329490614537873>
- Dawson, P. (2020). *Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education*. Routledge. <https://doi.org/10.4324/9780429324178>
- Dicks, A. P., Lautens, M., Koroluk, K. J. & Skonieczny, S. (2012). Undergraduate oral examinations in a university organic chemistry curriculum. *Journal of Chemical Education*, 89(12), 1506-1510. <https://doi.org/10.1021/ed200782c>
- Ellis, C., van Haeringen, K., Harper, R., Bretag, T., Zucker, I., McBride, S., Rozenberg, P., Newton, P. & Saddiqui, S. (2020). Does authentic assessment assure academic integrity? Evidence from contract cheating data. *Higher Education Research & Development*, 39(3), 454-469. <https://doi.org/10.1080/07294360.2019.1680956>
- Giordano, A. N. & Christopher, C. R. (2020). Repurposing best teaching practices for remote learning environments: Chemistry in the news and oral examinations during COVID-19. *Journal of Chemical Education*, 97(9), 2815-2818. <https://doi.org/10.1021/acs.jchemed.0c00753>
- Hart, C., Hammer, S., Collins, P. & Chardon, T. (2011). The real deal: Using authentic assessment to promote student engagement in the first and second years of a regional law program. *Legal Education Review*, 21(1/2), 97-121. <https://ler.scholasticahq.com/article/6244-the-real-deal-using-authentic-assessment-to-promote-student-engagement-in-the-first-and-second-years-of-a-regional-law-program.pdf>

- Hazen, H. (2020). Use of oral examinations to assess student learning in the social sciences. *Journal of Geography in Higher Education*, 44(4), 592-607. <https://doi.org/10.1080/03098265.2020.1773418>
- Huxham, M., Campbell, F. & Westwood, J. (2012). Oral versus written assessments: A test of student performance and attitudes. *Assessment & Evaluation in Higher Education*, 37(1), 125-136. <https://doi.org/10.1080/02602938.2010.515012>
- Iannone, P., Czichowsky, C. & Ruf, J. (2020). The impact of high stakes oral performance assessment on students' approaches to learning: A case study. *Educational Studies in Mathematics*, 103, 313-337. <https://link.springer.com/article/10.1007/s10649-020-09937-4>
- Iannone, P. & Simpson, A. (2012). Oral assessment in mathematics: Implementation and outcomes. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 31(4), 179-190. <https://doi.org/10.1093/teamat/hrs012>
- Iannone, P. & Simpson, A. (2015). Students' views of oral performance assessment in mathematics: Straddling the 'assessment of' and 'assessment for' learning divide. *Assessment & Evaluation in Higher Education*, 40(7), 971-987. <https://doi.org/10.1080/02602938.2014.961124>
- James, L. T. & Casidy, R. (2018). Authentic assessment in business education: Its effects on student satisfaction and promoting behaviour. *Studies in Higher Education*, 43(3), 401-415. <https://doi.org/10.1080/03075079.2016.1165659>
- Joughin, G. (1998). Dimensions of oral assessment. *Assessment & Evaluation in Higher Education*, 23(4), 367-378. <https://doi.org/10.1080/0260293980230404>
- Kang, D., Goico, S., Ghanbari, S., Bennallack, K. C., Pontes, T., O'Brien, D. H., & Hargis, J. (2019). Providing an Oral Examination as an Authentic Assessment in a Large Section, Undergraduate Diversity Class. *International Journal for the Scholarship of Teaching and Learning*, 13, (2), 1-14. <http://dx.doi.org/10.20429/ijso.tl.2019.130210>
- Knapp, T. R. & Mueller, R. O. (2010). Reliability and validity of instruments. In G. R. Hancock, L. M. Stapleton & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences*, 337-342. [2nd ed.] <https://www.routledge.com/The-Reviewers-Guide-to-Quantitative-Methods-in-the-Social-Sciences/Hancock-Stapleton-Mueller/p/book/9781138800137>
- Koh, S. B. K., Tai, S. Y. & Fung, F. M. (2021). Adopting a gamified approach of conducting viva voce in an undergraduate lab module. *Journal of Chemical Education*, 98(6), 2017-2022. <https://doi.org/10.1021/acs.jchemed.0c01397>
- Luckie, D. B., Rivkin, A. M., Aubry, J. R., Marengo, B. J., Creech, L. R. & Sweeder, R. D. (2013). Verbal final exam in introductory biology yields gains in student content knowledge and longitudinal performance. *CBE—Life Sciences Education*, 12(3), 515-529. <https://doi.org/10.1187/cbe.12-04-0050>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88, 105906. <https://doi.org/10.1016/j.ijssu.2021.105906>

- Pearce, G. & Lee, G. (2009). Viva voce (oral examination) as an assessment method: Insights from marketing students. *Journal of Marketing Education*, 31(2), 120-130. <https://doi.org/10.1177/0273475309334050>
- Pereira, D., Flores, M. A. & Niklasson, L. (2016). Assessment revisited: A review of research in *Assessment and Evaluation in Higher Education*. *Assessment & Evaluation in Higher Education*, 41(7), 1008-1032. <https://doi.org/10.1080/02602938.2015.1055233>
- Salamonson, Y., Metcalfe, L., Alexandrou, E., Cotton, A. H., McNally, S., Murphy, J. & Frost, S. A. (2016). Measuring final-year nursing students' satisfaction with the viva assessment. *Nurse Education in Practice*, 16(1), 91-96. <https://doi.org/10.1016/j.nepr.2015.09.003>
- Simper, T. (2010). A comparison of an oral assessment with a traditional paper exam within a final year nutrition module. *Educational Research and Reviews*, 5(8), 427-431. https://academicjournals.org/article/article1379621654_Simper.pdf
- Sotiriadou, P., Logan, D., Daly, A. & Guest, R. (2020). The role of authentic assessment to preserve academic integrity and promote skill development and employability. *Studies in Higher Education*, 45(11), 2132-2148. <https://doi.org/10.1080/03075079.2019.1582015>
- Sutherland, R. M., Reid, K. J., Chiavaroli, N. G., Smallwood, D. & McColl, G. J. (2019). Assessing diagnostic reasoning using a standardized case-based discussion. *Journal of Medical Education and Curricular Development*, 6. <https://doi.org/10.1177/2382120519849411>
- Tertiary Education Quality and Standards Agency (2015). *Higher Education Standards Framework (Threshold Standards) 2015*. Legislative instrument F2015L01639. Tertiary Education Quality and Standards Agency. Canberra: Australian Government. <https://www.legislation.gov.au/F2015L01639/latest/text>
- Thomas, J., Raynor, M. & McKinnon, M. (2014). Academic integrity and oral examination: An Arabian Gulf perspective. *Innovations in Education and Teaching International*, 51(5), 533-543. <https://doi.org/10.1080/14703297.2013.796724>
- Turjamaa, R., Hynynen, M. A., Mikkonen, I. & Ylinen, E. R. (2018). Dialogic oral exam in nursing education: A qualitative study of nursing students' perceptions. *Nurse Education in Practice*, 29, 53-58. <https://doi.org/10.1016/j.nepr.2017.11.008>
- Turner, K., Roberts, L., Heal, C. & Wright, L. (2013). Oral presentation as a form of summative assessment in a master's level PGCE module: The student perspective. *Assessment & Evaluation in Higher Education*, 38(6), 662-673. <https://doi.org/10.1080/02602938.2012.680016>
- Turner, M. & Davila-Ross, M. (2015). Using oral exams to assess psychological literacy: The final year research project interview. *Psychology Teaching Review*, 21(2), 48-68. <https://files.eric.ed.gov/fulltext/EJ1146560.pdf>
- 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(5), 840-854. <https://doi.org/10.1080/02602938.2017.1412396>
- Wallace, C. J. (2010). Using oral assessment in law: Opportunities and challenges. *The Law Teacher*, 44(3), 365-377. <https://doi.org/10.1080/03069400.2010.524035>

- Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation*, 2(1), article 2. <https://doi.org/10.7275/ffb1-mm19>
- Wiggins, G. (2011). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 92(7), 81-93. <https://doi.org/10.1177/003172171109200721>

Dr Shashi Nallaya (corresponding author) is a Senior Lecturer, Academic Development, in the Teaching Innovation Unit, University of South Australia, Adelaide, Australia. She holds a PhD from the University of Adelaide and MEd (TESL), BEd (TESL) and Dip. ESL degrees from Universiti Malaya, Malaysia.
ORCID: <https://orcid.org/0000-0003-3098-8112>
Email: shashi.nallaya@unisa.edu.au

Associate Professor Sheridan Gentili is the Director, Teaching Innovation Unit, University of South Australia, which is a central teaching and learning unit that partners and supports UniSA staff and external stakeholders to design and deliver an outstanding curriculum using relevant, effective teaching practices, technologies and tools.
ORCID: <https://orcid.org/0000-0001-5568-9106>
Email: sheridan.gentili@unisa.edu.au

Dr Scott Weeks is a Lecturer in Occupational Therapy in UniSA Allied Health and Human Performance, University of South Australia, Adelaide, Australia. He holds a PhD and a Bachelor of Applied Science (Occupational Therapy) with Honours from the University of South Australia. He focuses on shaping community wellbeing and educates future health professionals.
ORCID: <https://orcid.org/0000-0002-0795-2026>
Email: scott.weeks@unisa.edu.au

Associate Professor Katherine Baldock is the Dean of Programs (Human Performance), Allied Health and Human Performance, University of South Australia, Adelaide, Australia. She is a leader in higher education, including reimagining structural and cultural facilitators to support the transformative education of graduates who are socially just, anti-racist, culturally responsive, contributing to creating a better world.
ORCID: <https://orcid.org/0000-0003-3527-1057>
Email: Katherine.Baldock@unisa.edu.au

Please cite as: Nallaya, S., Gentili, S., Weeks, S. & Baldock, K. (2024). The validity, reliability, academic integrity and integration of oral assessments in higher education: A systematic review. *Issues in Educational Research*, 34(2), 629-646.
<http://www.iier.org.au/iier34/nallaya.pdf>