


# *Applying Kane's Validity Framework to Online OSCEs.*

Duana Quigley 

Julie Regan 

Trinity College, Dublin, [quigledu@tcd.ie](mailto:quigledu@tcd.ie)

## **Abstract.**

The methods we use to assess healthcare students' clinical ability, and justify our evaluations, have come to the fore in the context of competency-based healthcare education. The emphasis on examining the validity of student assessments more robustly is particularly relevant following the rapid transition to novel online examinations because of campus closures due to COVID-19 and repeated calls from many students and educators for online assessments to continue long-term. This paper describes the design and development of two online Objective Structured Clinical Examinations (OSCEs) and the application of Kane's (2013) established validity framework to the online OSCE for an undergraduate speech and language therapy programme. Assessment claims were produced, and evidence was gathered to rationalise these claims to generate a validity argument, which identified strengths and gaps that need to be addressed for future online OSCEs. The description of the process in this paper provides a theoretical and practical template for producing a validity argument for online OSCEs across a range of healthcare disciplines, and indeed for other common student assessment methods. Traditionally, a frequent method reported for deciding the value of assessments is solely capturing the perceptions of students and educators, which overlooks many of the other theoretical aspects of validity. This in-depth focus on producing a validity argument can enable educators to make a more objective, structured, holistic, and critical decision about whether the intended uses of student grades achieved from their chosen assessment can be defended.

**Keywords:** COVID-19; Medical education; Online assessment; OSCE; Validity.

## **1. Introduction.**

Within healthcare education, the learning objectives of many university programmes centre on ensuring students acquire the standards of proficiency required for their chosen profession and

consequently develop into safe and competent practitioners (CORU, 2014). Standards of proficiency are described as a triad of: (1) knowledge, such as theoretical concepts and practical principles; (2) skills, such as problem-solving and communication skills; and (3) other intangible elements, including values and attitudes (Sheepway, Lincoln & Togher, 2014). While many standards of proficiency are concrete and directly measurable, assessment of other competencies relies on inferences to be made on how students will act when they encounter specific clinical scenarios (Kane, 1999).

As there is a global trend to deliver competency-based healthcare education whereby assessment is aligned to clear learning outcomes, the decision-making processes associated with how we evaluate students' competencies has come to the fore (Boursicot et al., 2020). Therefore, the benefits and barriers of assessment processes need to be evaluated to ensure sound judgements are being made by educators in university programmes (i.e., the validity of assessments needs to be examined) (Cook, Brydges, Ginsburg & Hatala, 2015). The call to evaluate the validity of assessments of students' competency is particularly relevant to the implementation of online OSCEs across numerous healthcare education programmes who are bound by professional regulatory standards, including medicine, nursing, dental science, physiotherapy, occupational therapy, radiation therapy and speech and language therapy, amongst others.

### **1.1 What is an Objective Structured Clinical Examination (OSCE)?**

The Objective Structured Clinical Examination (OSCE) was originally developed in Dundee in 1972 as an assessment of students' clinical competence and skills (Harden, 1975). It is an assessment based on the principles of objectivity and standardisation and is used widely in the assessment of medical and healthcare students (Baid, 2011; Robbins & Hoke, 2008). Students rotate through a series of time-limited simulated assessment stations in a circuit. At each station students are examined using standardised grading rubrics by trained examiners. Khan, Gaunt, Ramachandran & Pushkar (2013) describe four different types of OSCE stations which enable a wide spread of clinical skills to be assessed: (i) observed station (i.e., examiner is present for the duration of testing); (ii) unobserved station (i.e., no examiner is present and answers are submitted on paper after the station); (iii) technology enhanced station (i.e., station involving the use of modern technology such as manikins) and (iv) linked stations (i.e., two consecutive stations that refer to the same clinical case or information).

Documented advantages of OSCEs are plentiful in the literature, such as increased student confidence, better preparation for clinical practice, and augmented self-awareness of clinical strengths and areas to develop (Barry, Noonan, Bradshaw & Murphy-Tighe, 2012; Ghouri et al., 2018). A study of student speech and language therapists' (SLTs) perceptions of OSCEs reported similar positive findings, indicating that they considered the OSCE easy to understand, fair, representative of clinical practice, confidence-boosting, and a meaningful method of assessing their clinical competencies and highlighting areas to develop (Quigley & Regan, 2021). Conversely, potential disadvantages of OSCEs have also been identified such as intensity of resources required in terms of staff, space, and equipment (Baid, 2011). Increased student stress associated with OSCEs has also been documented (Fidment, 2012). Likewise, over one third of student SLTs believed the OSCE was more stressful than written exams (Quigley & Regan, 2021).

## **1.2 COVID-19 and Introduction of Online Assessments.**

Coronavirus disease 2019 (COVID-19) was categorised as a pandemic by the World Health Organization (WHO) and led to widespread introduction of public health measures globally that impacted on third level education, such as campus closures and online teaching and assessment (Choi et al., 2020). Consequently, many OSCEs that were traditionally assessed in a face-to-face manner were implemented online out of necessity due to the COVID-19 pandemic (e.g., Boyle et al., 2020; Craig, Kasana & Modi, 2020; Hannan, Umar, Rob & Choudhury, 2021; Hannon, Lappe, Griffin & Roussel, 2020; Kakadia, Chen & Ohyama, 2020). Initial guidance was available from educators who had previously completed OSCEs online pre-COVID, and who documented positive student experiences and many similarities with face-to-face OSCEs (e.g., Langenau, Kachur & Horber, 2014; Prettyman, Knight & Allison, 2018).

Favourable feedback reported by examiners and students indicated that the online OSCEs were comparable to traditional OSCEs (Craig et al., 2020; Kakadia et al., 2020; Lara, Foster, Hawks & Montgomery, 2020; Palmer et al., 2015), with some participants going as far as describing them as enjoyable and enhanced student performance (Langenau et al., 2014; Prettyman et al., 2021). Key advantages included their flexibility and convenience, as they enabled the ability to assess a range of students' clinical skills from a distance through telehealth stations, thereby reducing travel time and associated costs while adhering to physical distancing guidelines (Palmer et al., 2015; Prettyman et al., 2021; Shehata et al., 2020). Online OSCEs typically require no additional equipment or software as they can be administered using low-cost or free

platforms that students and examiners are already familiar with (Prettyman et al., 2021). In addition, online OSCEs have been found to be successful in enabling a summative assessment of clinical competence (Boyle et al., 2020) and were perceived by examiners to be a valid assessment method (Chan, Humphrey-Munro, Pugh, Su & Wood, 2014). However, some examiners reported barriers using online OSCEs to accurately complete physical examinations and therefore raised a concern that online OSCEs only assess the '*knows how*' level of Miller's pyramid rather than the '*shows how*' level for this clinical skill (Blythe et al., 2021; Craig et al., 2020; Hannon et al., 2020). Challenges were also raised in relation to IT issues, such as internet connection instability and potential data protection concerns when using online platforms (Boyle et al., 2020; Hannan et al., 2021; Kakadia et al., 2020). Some students reported difficulties making inter-personal connections and demonstrating empathy in the online interactions with examiners/simulated patients (Hannon et al., 2020; Langenau et al., 2014). Examiners highlighted the increased preparation and examiner training required (Hannan et al., 2021; Shehata et al., 2020).

Despite the documented challenges, many researchers concluded that online OSCEs are a feasible assessment method in healthcare education, but there have been calls for their validity to be evaluated more robustly. Some argue that existing OSCE validity research is somewhat narrow, focusing too much on the validity of the content of the OSCE station or the validity of the OSCE process, without considering the possible impact of broader factors (Hodges, 2003). This concern with validity is particularly relevant to the rapid introduction of online OSCEs in the context of online teaching and assessment during the COVID-19 pandemic, which may continue long-term into the future.

### **1.3 Evaluating the Validity of Online Assessments.**

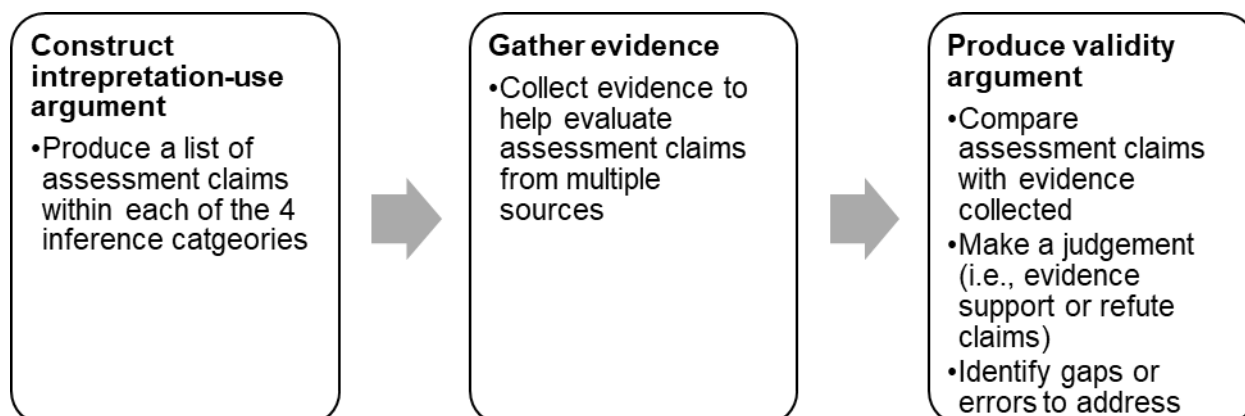
The validity of an assessment is an evaluation of how plausible the assessment claims are based on the scores achieved (Kane, 2013). For example, assessment claims in healthcare education may lead to an interpretation of a student's level of competence, may justify a student's selection for a particular clinical placement, or may support evaluation of programme learning outcomes. Traditionally, evaluation of validity relied on three measures: content validity (i.e., reliability of content of the assessment items), criterion validity (i.e., correlation with another standardised measure of the same item) and construct validity (i.e., intangible constructs matched to observable behaviours) (Cook et al., 2015). Similarly, a discourse analysis of diverse

ways in which validity is described found that validity is construed by some as an indicator that an assessment has achieved a gold standard, while considered by others to accurately interpret scores or the value of assessments for students and society (St. Onge, Young, Eva & Hodges, 2017). Contemporary methods of measuring validity of assessments within healthcare education tend to favour Kane's validity framework due to its versatility and widespread applicability (e.g., Cook et al., 2015; Hess & Kvern, 2021). Kane (2013) recommends that educators should produce a list of assessment claims, what he termed '*interpretation-use argument (IUA)*' and asserts that any such assessment claims should be supported by appropriate evidence to rationalise the decisions made based on the test scores. Kane's (2013) validity framework specifies four central inferences, each of which requires evidence, to help support a validity claim:

- scoring (i.e., providing a score or assessment claim for the observed competency, including how the data was gathered, recorded, and scored)
- generalisation (i.e., generalising the score from the assessment environment to expected performance in a different environment or how similar it would be if the student was retested)
- extrapolation (i.e., extrapolating the scores from the assessment environment to the target real-world domain such as clinical practice)
- implications (i.e., drawing on the scores to assist decision-making in a manner that can be easily defended).

Hess and Kvern (2021) provide a useful list of examples of validity evidence for healthcare educators for each of the four inferences. Once evidence is gathered from multiple sources, a validity argument can be produced by comparing the original assumptions with the actual evidence collected, determining whether to accept/reject/revise the IUA and identifying any gaps or errors that may need to be addressed (Kane, 2013). Ultimately, this allows a decision to be made about whether the intended uses of student grades achieved from the assessment can be defended (Kane, 2013). This process of producing a validity argument is summarised in Figure 1.

**Figure 1. Producing a validity argument (based on (Kane, 2013)).**



Kane's (2013) validity framework has been applied previously to online OSCEs with paramedic students (Tavares et al., 2018) and assessment of the professionalism of medical student based on observed behaviours (Clauser, Margolis, Holtman, Katsufrakis & Hawkins, 2012). It was reported Kane's (2013) validity framework assisted with structuring, organising, and conceptualising the inter-relationships between hypotheses, analyses, and interpretations. A further strength of Kane's (2013) validity framework highlighted by Schuwirth and van der Vleuten (2012) is its applicability to qualitative assessments as well as psychometric data.

The aims of this study were: (i) to design and develop two online OSCEs to assess the clinical competencies of student speech and language therapists; (ii) to gather perspectives of students and educators about the online OSCEs; (iii) to apply Kane's (2013) validity framework to the online OSCEs to determine how plausible the assessment claims are; and (iv) to provide recommendations for designing and administering assessment of clinical competencies through online OSCEs in the future, based on the validity argument produced.

## 2. Methods.

In the section, we describe the setting of the study and the online OSCE design, outline our assessment claims and the evidence we gathered to test these claims to produce the validity argument.

### 2.1 Setting.

The study was conducted in Trinity College Dublin where a four year full-time undergraduate programme for student speech and language therapists is delivered. OSCEs are administered

at two time points within the speech and language therapy programme:

(i) Year 2 in advance of the students' first clinical placement with a focus on the clinical competencies required to assess the communication and swallowing abilities of children and adults. This OSCE assesses clinical skills taught during a 12-week Practice Education module in term 1, which aims to translate academic theory to clinical practice for the assessment of children and adults presenting with a range of speech, language, communication, and swallowing disorders.

(ii) Year 4 to assess students' clinical skills within the specific clinical context of dysphagia. This OSCE assesses students' clinical competence taught during a 12-week module in term 1 on the assessment, differential diagnosis and management of clients presenting with swallowing disorders across a vast spectrum of aetiologies and co-morbid medical or neurological conditions.

Previously, OSCEs were administered face-to-face, but in the academic year 2020/2021 during the COVID-19 pandemic, the OSCEs were implemented fully online. The OSCEs were administered at the end of each 12-week module (end of term 1).

## **2.2 Online OSCE Design.**

The content of the online OSCEs were designed with consideration of the learning outcomes of the programme and were blueprinted using the Irish national student clinical competency performance indicators. The aim of this examination was that the assessment scores from the online OSCEs would inform decisions regarding a student's level of competence in advance of clinical placement.

The authors consulted the principles of OSCE design and administration set out by Khan et al. (2013). We also determined what adaptations and additions were needed for an online modality in the planning, implementation, and evaluation phases (Shehata et al., 2020). For example, considered what IT platform and functions to use and data protection implications, changed the type and sequence of stations, adapted the grading rubric, clarified the role of each staff member, provided additional examiner training, and planned how to evaluate the process. In addition, Daniels and Pugh's (2018) twelve tips for developing an OSCE, which are informed by Kane's validity framework, provided a helpful guide for designing the online OSCEs. OSCE scenarios were developed based on practitioners' work experiences, typical clinical scenarios within the SLT scope of practice, and aligned with performance expectations.

Each online OSCE comprised a circuit of 6-8 online stations, that took place on the IT platform Zoom™ or Blackboard Collaborate Ultra™. At each online station, students were individually assessed by an examiner whilst completing a specific clinical task, such as administration of a standardised test, taking a case-history, analysing clinical data or selecting treatment goals and procedures (see Figure 2 for an example of an OSCE circuit for year 2 students). The majority of OSCE stations fell within Khan et al.'s (2013) category of “*observed station*”, that is, an examiner was present for the duration of the OSCE station assessment, while one OSCE station in each year was a “*technology enhanced station*”. For year 2, the technology enhanced station involved watching a video of a healthcare interaction and reporting the findings and for year 4 it required students to interpret the result of a videofluoroscopy examination (i.e., interpret a moving x-ray of a patient’s swallow). The stations were ten minutes duration and students had a five-minute break between stations. A super host, who is a member of the teaching staff, acted as a nominated point of communication for students and examiners before, during and after the OSCE. This individual coordinated and oversaw the entire examination process and were available to trouble shoot any issues if they arose, to ensure a smooth and efficient exam experience for students and examiners.

**Figure 2. Online OSCE station circuit for year 2 students.**





## 2.3 Examination and Grading.

Each OSCE station was assessed by one examiner, who was a qualified SLT and was a member of the academic or clinical teaching staff. All examiners were provided with an orientation and training session which described the OSCE stations, role of the examiner within their station, marking rubric and how to complete it. The marking rubrics and pass criteria were rooted in clinical decision making and the performance expected by students as outlined in the national student clinical competency evaluation framework. Students' grades were based on the combined score of all stations. The year 2 OSCE grade represented 30% of the total module grade, while the year 4 OSCE graded represented 100% of the module grade.

On the day of the online OSCE, examiners admitted individual students to the online station, while their peers remained in the virtual waiting room. Based on their observations of a student's live performance and answers to the questions they posed, the examiners entered grading scores and any relevant narrative comment for each student on an electronic excel document for later analysis. Students were provided with written, individual feedback via a personalised email within one week of the online OSCE completion that outlined their strengths and areas to develop based on examiners' observations.

## 2.4 Construction of Assessment Claims to Help Produce the Validity Argument.

Drawing on the process outlined in Figure 1 and the recommendations of Hess and Kvern (2021) and Clauser et al. (2012), we documented our assumptions and produced a list of assessment claims for the online OSCEs in each of Kane's (2013) four inference categories: scoring, generalization, extrapolation, and implications as follows:

Scoring:

- The year 2 OSCE will accurately test competence in clinical assessment skills prior to first placement.
- The year 4 OSCE will accurately test competence in clinical skills in a discrete area of practice at graduate level (i.e., dysphagia).

Generalization:

- Sampling of relevant skills will be adequate to establish an assessment of students' competence.

- Grades achieved in the assessment environment will be comparable if different clinical scenarios are presented.

Extrapolation:

- Students' performance on this assessment will predict performance of a range of competencies in the clinical learning environment while the student is on placement.
- Scope of OSCE reflects SLT scope of practice.

Implications:

- Students will be provided with information about their clinical strengths and areas to develop.
- Students who fail will be identified for supportive remediation.
- Students who pass the OSCE are more likely assumed to pass their clinical placement.
- Students and examiners will perceive the online OSCE positively.

## **2.5 Collection of Evidence to Help Evaluate Assessment Claims to Produce a Validity Argument.**

We collected the evidence available from multiple sources to support (or refute) each inference. Evidence collected included pre-training content for OSCE examiners, OSCE schedule, OSCE exam paper, marking rubrics, student scores, and feedback sheets to students which included student scores and examiners' qualitative comments (see full list of evidence in Table 1). As part of the gathering of evidence, ethical approval was also obtained from the Research Ethics Committee within the university to evaluate student and examiner perspectives of the online OSCE (Ethical approval reference: MT11). A prospective cross-sectional study using an online survey design was conducted. To help inform the survey design and content, we consulted with previous studies that evaluated perspectives of OSCEs using a survey tool (Barry et al., 2012; Graham, Zubiaurre Bitzer & Anderson, 2013). We also generated our own bespoke survey questions that related to the type of OSCE stations implemented and the process completed. The survey was piloted prior to its distribution with two individuals not otherwise involved in the study, to help ensure the questions and design of the survey were user-friendly and accessible. Within one week of both OSCEs, the online anonymous survey was disseminated to the participating undergraduate SLT students and OSCE examiners. The survey collected quantitative data (e.g., responses to 10 Likert scale questions) and qualitative data (e.g., free

field text). Quantitative data were analysed both descriptively and using Fisher's exact tests. Thematic analysis was used to analyse free field comments.

Next, we compared our assessment claims with the evidence collected, as will be discussed below. This enabled us to make a judgement on the validity argument and identify gaps or errors to address for future online OSCEs.

### **3. Results.**

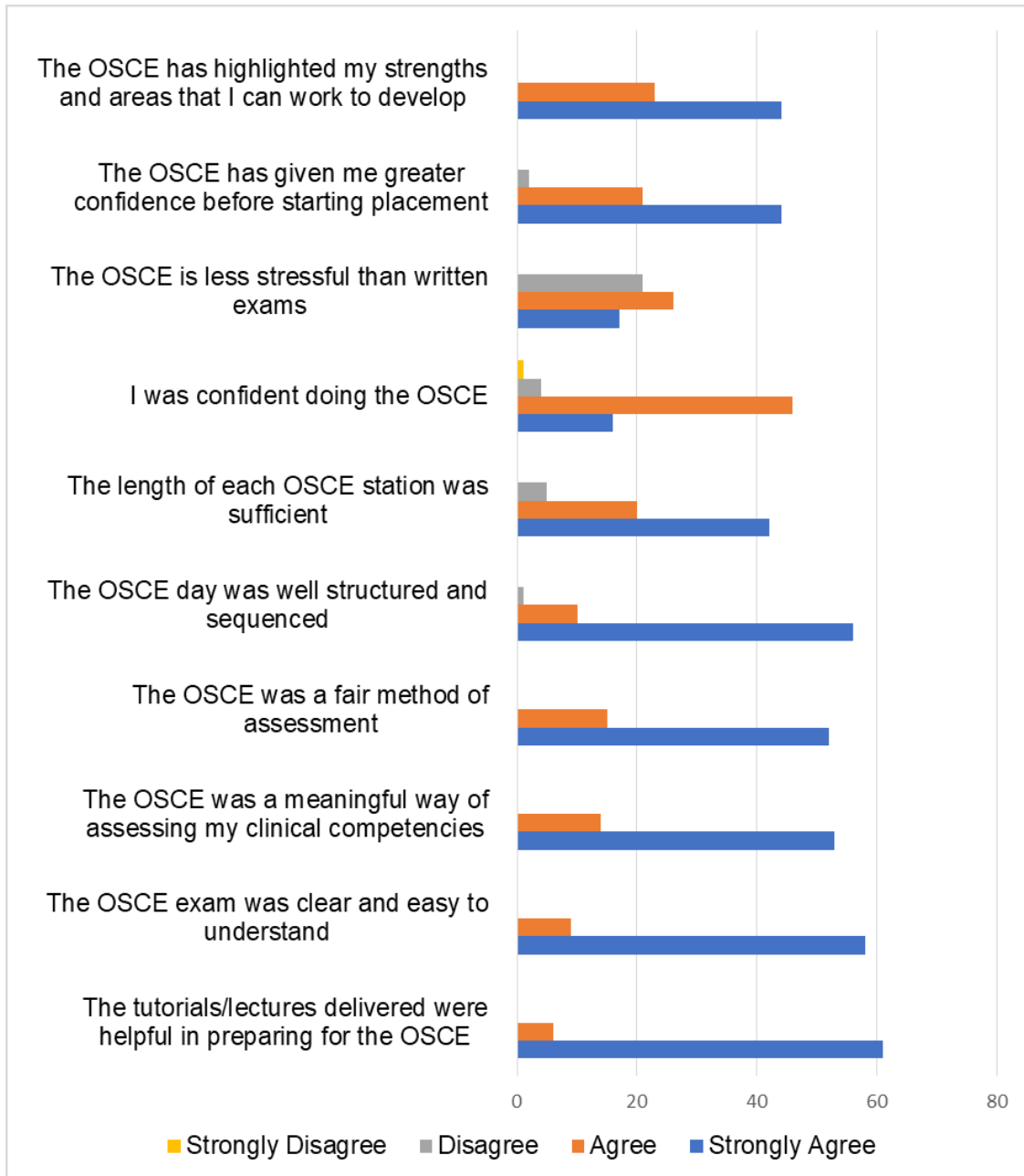
A total of 70 students completed the online OSCE and 10 qualified SLTs acted as examiners. 34 students were at year 4 level and 36 students were at year 2 level.

#### **3.1 Results of Survey of Students' and Examiners' Perspectives.**

67 students (96% response rate) and 10 examiners (100% response rate) responded to the anonymous survey. As outlined in Figure 3, students rated the online OSCE favourably overall, with the majority of students strongly agreeing or agreeing with statements related to its length, structure, sequence, fairness, meaningfulness and ability to instil confidence and highlight strengths and areas to develop.

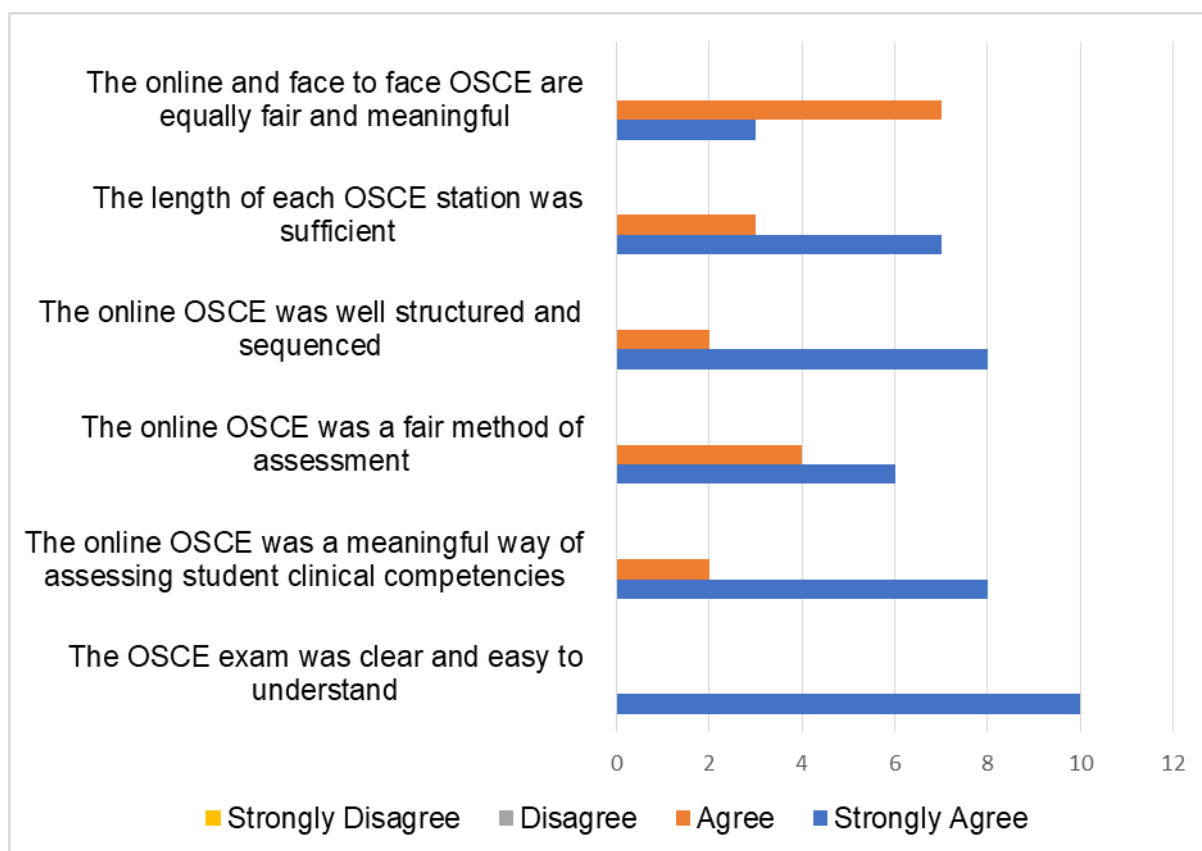
Thematic analysis of free field comments generated four key themes that highlighted advantages and disadvantages of the online OSCEs. Advantages included preparation for clinical practice, practise using telehealth, time efficient, fairness, adherence to public health guidance and infection control protocols, ability to complete from home. Disadvantages included stress-inducing, worries and experiences of IT connectivity issues and a perception that it was not suitable to assess all clinical competencies. The year 4 students who had previously completed face-to-face OSCE when they were in year 2 (n=33) were asked to indicate which format they preferred: 64% preferred the online format and 100% agreed or strongly agreed that the online and face-to-face OSCE are equally fair and meaningful.

**Figure 3. Student SLTs' perceptions of online OSCEs.**



Likewise, examiners rated the online OSCE positively, with no examiner disagreeing or strongly disagreeing with any statement (Figure 4). Examiners echoed many of the same advantages and disadvantages identified by students. They also commended the efficiency of online grading but noted increased examiner demands managing the IT connectivity and access of each student to the IT platform.

**Figure 4. Examiners' perceptions of online.**



### 3.2 Validity Argument Produced.

Based on the process outlined in Figure 1, we produced a validity argument for the online OSCEs that outlines which assumptions and assessment claims were sufficiently supported and which held inadequate evidence and require further action. This is summarised in Table 1.

**Table 1. Validity argument produced for online OSCEs in undergraduate speech and language therapy programme.**

Inference category and description from Kane's (2013) Validity Framework	Assessment Claims	Evidence gathered to support assessment claims	Validity argument
<b>Scoring - score or</b>	The year 2 OSCE will	Scoring of observed behaviours was	Moderate support of assessment claims

<p><i>assessment claim for the observed competency.</i></p>	<p>accurately test competence in clinical assessment skills prior to first placement</p> <p>The year 4 OSCE will accurately test competence in clinical skills in a discrete area of practice at graduate level (i.e., dysphagia)</p>	<p>standardised in terms of clinical scenario, duration of observation and setting. Examiners had the option of providing a free field comment also about observed behaviour.</p> <p>Marking rubric provided clear criteria and was developed with the input of experienced clinicians.</p> <p>Examiners were experienced practitioners working in healthcare environments and were familiar with the competencies required and the online platform used.</p> <p>Examiners were provided with an orientation and training session about the OSCE stations, role of the examiner within their station, marking rubric and how to complete it.</p> <p>Assessment of each competency was supported by observation of performance via videoconferencing in real time by the same examiner.</p> <p>A super host was nominated to support quality assurance and to ensure any IT issues were addressed and time was</p>	<p>evident and recommendation to address the following gaps:</p> <ul style="list-style-type: none"> <li>• Further reliability analyses required such as intra- and inter-rater reliability</li> <li>• Analysis of the distribution of ratings provided by each rater to be completed</li> <li>• Think-aloud protocols that capture the examiner's thought process when completing the grading rubric</li> <li>• Direct observation training of examiners</li> <li>• Standard setting e.g., Borderline Regression Method</li> </ul>
---	---	---	--

		scheduled in between each station.	
<p><b>Generalisation</b> -Generalisation of the score from the assessment environment to expected performance in a different environment.</p>	<p>Sampling of relevant skills will be adequate to establish an assessment of students' competence.</p> <p>Grades achieved in the assessment environment will be comparable if different clinical scenarios were presented.</p>	<p>Clinical competencies assessed were drawn from the blueprint of national clinical competency performance indicators to ensure adequate sampling.</p> <p>Students were assessed by multiple examiners (different examiner in each station) to maximise sampling of skills by different observers, different domains of practice and different data types.</p> <p>Student's grade was calculated based on total combined score.</p>	<p>Moderate support of assessment claims evident and recommendation to address the following gaps:</p> <ul style="list-style-type: none"> <li>• Evaluation of whether performance online predicts performance in a different face-to-face environment</li> <li>• Evaluation of whether performance rating was influenced by other factors</li> <li>• Randomized controlled trials to examine the comparability of the performance scores, and their generalization, between face-to-face and online OSCEs</li> <li>• Test assessment claims with larger sample size</li> </ul>
<p><b>Extrapolation</b> -Extrapolation of the score from the assessment environment to the target real-world domain.</p>	<p>Students' performance on this assessment will predict performance of a range of competencies in the clinical learning environment while the student is on placement.</p> <p>Scope of OSCE</p>	<p>Experienced clinicians helped design the OSCE stations to ensure their appropriateness and relevance to real-world clinical practice.</p> <p>In line with public health and clinical guidelines, no assessment of clinical skills was completed that were aerosol generating or required physical manipulation.</p>	<p>Weak support of assessment claims evident and recommendation to address the following gaps:</p> <ul style="list-style-type: none"> <li>• Correlations between OSCE result and placement result</li> <li>• Consideration of the possible impact of reactivity bias via long-term observations in numerous settings</li> </ul>

	<p>reflects SLT scope of practice.</p>		<ul style="list-style-type: none"> <li>• Correlations between OSCE result and academic exam result</li> <li>• Correlations between OSCE result and perceptions of patients, healthcare managers, administrators, or policy makers</li> <li>• Evaluation of extrapolation of scores to clinical skills that are aerosol generating or require physical manipulation</li> <li>• Evaluation of differences in decision-making completed in online OSCEs</li> <li>• Evidence that scores improve following remediation training</li> </ul>
<p><b>Implications</b> - Drawing on the scores to assist decision-making</p>	<p>Students will be provided with information about their clinical strengths and areas to develop.</p> <p>Students who fail will be identified for supportive remediation.</p> <p>Students who pass the OSCE will pass their clinical placement.</p>	<p>Individual student feedback was provided after each OSCE.</p> <p>No competency expectations were modified to fit the virtual format.</p> <p>Online OSCEs did not change the curriculum content or delivery.</p> <p>Perspectives of students and examiners were overwhelmingly positive on most aspects, with disadvantages identified relating to IT connectivity, student stress and not being suitable for</p>	<p>Moderate support of assessment claims evident and recommendation to address the following gaps:</p> <ul style="list-style-type: none"> <li>• Comparison of pass-fail results from online OSCEs with face-to-face OSCEs</li> <li>• Evaluation of effectiveness of remediation measures provided</li> <li>• Evaluation of intended and unintended consequences of the OSCE (e.g., quality of feedback received).</li> </ul>



	Students and examiners will perceive the online OSCE positively.	assessing all clinical competencies (see Figure 3 and 4).	
--	--	---	--

## 4. Discussion.

This study aimed to apply Kane's (2013) validity framework to online OSCEs that assess the clinical competencies of student speech and language therapists, to help determine how plausible this format of assessment is and provide recommendations for the future. The validity argument produced (Table 1) documents strengths and weaknesses in the assumptions made and the inference chain and proposes recommendations to address identified gaps.

The analysis completed suggests the evidence gathered provides moderate support for the assessment claims related to scoring, generalisation and implication inferences. Strengths of the validity argument centre on the methods used to observe, collect, and document students' competency were appropriate and that the grading rules were applied and documented accurately (i.e., scoring). In addition, supportive evidence was available that grading rubrics were blueprinted against national performance indicators and sampling of skills was maximised by placing a different examiner in each station (i.e., generalisation). Furthermore, evidence was available to support the claim that no changes were made to the curriculum content or delivery, students were provided with constructive feedback about their performance, and that students and examiners rated the online OSCEs positively (i.e., implication). Nevertheless, gaps were identified and recommendations to address shortcomings are proposed to strengthen the former inferences as outlined in Table 1. Examiner variability has been recognised as a factor that may support a validity argument by some researchers as it is considered to reflect the reality of healthcare delivery, whereby practitioners will address the complexity and multiple factors of clinical practice in different ways (Boursicot et al., 2020). However, others suggest that when scoring relies on human judgment, as it did in the online OSCEs in this study, resulting examiner variability may reduce the generalizability of the score (Clauser et al., 2012). Therefore, additional supportive evidence may be gathered through think-aloud protocols that capture the examiner's thought process when completing the grading rubric or analysis of the stability of grades across stations or raters, such as coefficient kappa (Brennan, 2001; Stern 1996). While all examiners were experienced clinicians and received orientation and training about the OSCE

stations, their role as examiner, marking rubric and how to complete it, the validity argument may be strengthened if direct observation training was provided to help increase inter-rater reliabilities (Holmboe, Hawkins & Huot, 2004). Randomised controlled trials may help compare students' performance in online OSCE vs face-to-face OSCE (Hess & Kvern, 2020). Furthermore, a larger sample of online OSCE stations, greater number of examiners and an increased number of items on the grading rubric would help support the generalisation assessment claim (Hatala et al., 2015). Standard setting may also be supported by introducing the Borderline Regression Method for the discipline, a criterion-reference standard setting commonly used in assessment of medical students (McKinley & Norcini, 2014).

The evidence collected provides weak support for the assessment claims related to the extrapolation inference, that is, the ability to extrapolate the scores from the online OSCE to the real-world domain of SLT clinical practice. Although some evidence for content validity was collected (i.e., experienced clinicians helped design the OSCE stations to ensure their appropriateness and relevance to real-world clinical practice), this only supports what the educators' objective was. While some stress that not all inferences in the validity argument are of equal importance (Cook et al., 2015), others argue that the overall validity argument is only as robust as the weakest link in the chain of assumptions (Clauser et al., 2012). Kane (2013) stresses that the intended use of the assessment tool will determine which evidence should carry more weight (e.g., scoring and generalisation evidence for high stakes summative assessment to support standardisation vs. extrapolation evidence for formative assessments to help inform behaviours in clinical practice). As the online OSCEs in this study were intended to assess clinical competence in advance of placement, we recommend further research is conducted to gather evidence for the extrapolation inference which in turn can help support or refute the validity argument for online OSCEs. This would increase its potential capacity to reliably predict how students' clinical skills observable online transfer to the healthcare setting. For example, comparing the students' performance in the online OSCE stations and their performance in the real world of clinical practice whilst on placement, or when engaging in high-fidelity simulations, or another comparable assessment (Boursicot et al., 2020). In addition, gathering evidence as to whether students' grades would extrapolate to clinical skills not assessed in the online OSCEs due to public health guidance during the COVID-19 pandemic at the time, such as recommendations to reduce aerosol generating procedures or practices that require physical manipulation where possible. Hatala et al. (2015) also suggests collecting evidence that student scores improve following remediation training and evaluating how

feedback provided to students impacts on their subsequent performance. If educators were to conceptualise validity as a social imperative, judgements of validity in the extrapolation domain would be expanded to perceptions of patients, healthcare managers, administrators, or policy makers, instead of validity evidence resting solely with the properties of the actual assessment tool (St. Onge et al., 2017). Socio-cognitive theory may also provide insights into differences, if any, of decision-making completed in online OSCEs (Hess & Kvern, 2021).

This study provides theoretical and practical direction for healthcare educators on the principles, processes, and evidence to consider when applying Kane's validity framework to online OSCEs in their discipline. While three of the four inferences of the validity argument documented in this context received moderate support for the assessment claims, clear direction and suggestions for future iterations and evaluations of the online OSCE is provided to help augment the existing evidence and strengthen the validity claims. Many of these formal aspects of validity might be overlooked if relying on student and examiner perceptions only, which is a common method reported in the literature when deciding the value of assessments (Brennan, 2013). The importance of assessment validity and its numerous facets is equally applicable to a range of common student assessments in healthcare education across diverse disciplines, such as portfolio-based assessment and other qualitative examinations (Cook et al., 2015).

## 5. Conclusion.

Validity is often used to confirm the high quality of assessment tools and justify how decisions are made based on assessments of healthcare students. This study provides an example of applying Kane's (2013) validity framework to an online OSCE in the context of undergraduate speech and language therapy education which could be replicated by other disciplines. It enabled an objective, structured, holistic, critical reflection on the introduction of online OSCEs and identified future modifications and evaluations required to strengthen the evidence to demonstrate the validity of the online OSCE and justify the decisions that are made based on student scores. It is very likely post-COVID that both online and face-to-face OSCEs will be implemented due to their respective advantages and therefore it is timely to gather the necessary evidence to support the validity argument for online OSCEs and other distance assessments.

## 5.1 Limitations.

The validity argument we produced may have been negatively impacted by the relatively small sample size of students and examiners who participated in online OSCEs in a single discipline (Swanson, Clauser & Case, 1999). Students and examiners from other healthcare disciplines and based in other jurisdictions may report different results. We also acknowledge the potential subjectivity in constructing the validity argument (Table 1) for the online OSCE that we were involved in designing and evaluating.

## 5.2 Acknowledgements.

We would like to thank the students and examiners who participated in the online OSCEs described in this paper

## 6. References.

- Baid, H. (2011). The objective structured clinical examination within intensive care nursing education. *Nursing in Critical Care*, 16(2), 99–105.
- Barry, M., Noonan, M., Bradshaw, C., & Murphy-Tighe, S. (2012). An exploration of student midwives' experiences of the Objective Structured Clinical Examination assessment process. *Nurse Education Today*, 32(6), 690–694.
- Blythe, J., Patel, N., Spiring, W., Easton, G., Evans, D., Meskevicius-Sadler, E., Noshib, H., & Gordon, H. (2021). Undertaking a high-stakes virtual OSCE (“VOSCE”) during COVID-19. *BMC Medical Education*, 21: 221 <https://doi.org/10.1186/s12909-021-02660-5>
- Boursicot, K., Kemp, S., Wilkinson, T., Findyartini, A., Canning, C., Cilliers, F. & Fuller, R. (2020). Performance Assessment: Consensus statement and recommendations from the 2020 Ottawa Conference. *Medical Teacher*, 43(1), 58-67.
- Boyle, J., Colquhoun, I., Noonan, Z., McDowall, S., Walters, M. & Leach, J. (2020). Viva la VOSCE? *BMC Medical Education*, 20, 514 <https://doi.org/10.1186/s12909-020-02444-3>
- Brennan, R. L. (2001). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement*, 38(4), 295–317.
- Brennan, R.L. (2013). Commentary on “Validating the Interpretations and Uses of Test Scores”. *Journal of Educational Measurement*, 50(1), 74-83.  
<https://doi.org/10.1111/jedm.12001>
- Chan, J., Humphrey-Munro, S., Pugh, D., Su, C., & Wood, T. (2014). The objective structured

- clinical examination: can physician-examiners participate from a distance? *Medical Education*. 48(4), 441-450.
- Choi, B., Jegatheeswaran, L., Minocha, A., Alhilani, M., Nakhoul, M. & Mutengesa, E. (2020). The impact of the COVID-19 pandemic on final year medical students in the United Kingdom: a national survey. *BMC Medical Education*, 20 (206), <https://doi.org/10.1186/s12909-020-02117-1>
- Clauser, B., Margolis, M., Holtman, M., Katsufakis, P. & Hawkins, R. (2012). Validity considerations in the assessment of professionalism. *Advances in Health Science Education*, 17(2), 165-181.
- Cook, D., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*, 49(6), 560-575.
- CORU. (2014). *Speech and Language Therapist Registration Board. Criteria and Standards of Proficiency for Education and Training* (Dublin: CORU).
- Craig, C., Kasana, N., & Modi, A. (2020). Virtual OSCE delivery: The way of the future? *Medical Education Adaptations*, 54(12), 1185-1186.
- Daniels, V., & Pugh, D. (2018). Twelve tips for developing an OSCE that measures what you want. *Medical Teacher*, 40(12), 1208-1213.
- Fidment, S. (2012). The Objective Structured Clinical Examination (OSCE): A qualitative study exploring the healthcare student's experience. *Student Engagement & Experience Journal*, 1(1), 1-11.
- Ghouri, A., Boachie, C., McDowall, S., Parle, J., Ditchfield, C., McConnachie, A., Walters, M. & Ghouri, N. (2018). Gaining an advantage by sitting an OSCE after your peers. *Medical Teacher*, 40(11), 1136-1142.
- Graham, R., Zubiaurre Bitzer, L., & Anderson, O. (2013). Reliability and predictive validity of a comprehensive preclinical OSCE in dental education. *Journal of Dental Education*, 77(2), 161-167.
- Hannan, T., Umar, S., Rob, Z. & Choudhury, R. (2021). Designing and running an online Objective Structured Clinical Examination (OSCE) on Zoom: A peer-led example. *Medical Teacher*, 43(6), 651-655.
- Hannon, P., Lappe, K., Griffin, C., & Roussel, D. (2020). Objective structured clinical examination: from exam room to zoom breakout room. *Medical Education*, 54(9), 861-861.
- Harden, R., Stevenson, M., Downie, W., & Wilson, G. (1975). Assessment of clinical

- competence using objective structured examination. *British Medical Journal*, 22(1), 447-451.
- Hatala, R., Cook, D.A., Brydges, R., & Hawkins, R. (2015). Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): a systematic review of validity evidence. *Advances in Health Science Education*, 20(5), 1149–1175.
- Hess, B. & Kvern, B. (2020). A call to understand the psychometric implications of virtual OSCE delivery. *Medical Teacher*, 43(7), 860.
- Hess, B. & Kvern, B. (2021). Using Kane's framework to build a validity argument supporting (or not) virtual OSCEs. *Medical Teacher*, 43(9), 999-1004.
- Hodges, B. (2003). Validity and the OSCE. *Medical Teacher*. 25(3):250-254.
- Holmboe, E. S., Hawkins, R. E., & Huot, S. J. (2004). Effects of training in direct observation of medical residents' clinical competence: A randomized trial. *Annals of Internal Medicine*, 140(11), 874–881.
- Kakadia, R., Chen, E., & Ohyama, H. (2020). Implementing an online OSCE during the COVID-19 pandemic. *Advancing Through Innovation*, 85(S1), 1006-1008.
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Khan, K. Z., Gaunt, K., Ramachandran, S., & Pushkar, P. (2013). The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: organisation & administration. *Medical Teacher*, 35(9), e1447–e1463. <https://doi.org/10.3109/0142159X.2013.818635>
- Langenau, E., Kachur, E., Horber, D. (2014). Web based objective structured clinical examination with remote standardised patients and Skype: Resident experience. *Patient Education and Counselling*. 96(1), 55-62.
- Lara, S., Foster, C., Hawks, M., & Montgomery, M. (2020). Remote assessment of clinical skills during COVID-19: a virtual, high-stakes, summative pediatric objective structured clinical examination. *Academic Pediatrics*, 20(6),760–761.
- McKinley, D., & Norcini, J. (2014). How to set standards on performance-based examinations: AMEE Guide No. 85. *Medical Teacher*, 36(2), 97–110.
- Palmer, R., Biagioli, F., Mujcic, J., Schneider, B., Spires, L., & Dodson, L. (2015). The feasibility and acceptability of administering a telemedicine objective structured clinical exam as a solution for providing equivalent education to remote and rural learners. *Rural and Remote Health*. 15(4): 3399.
- Prettyman, A., Knight, E., Allison, T. (2018). Objective structured clinical examination from

- virtually anywhere! *The Journal for Nurse Practitioners*, 14(8), e157-e163.
- Quigley, D., & Regan, J. (2021). Introduction of the Objective Structured Clinical Examination in Speech and Language Therapy Education: Student Perspectives. *Folia Phoniatrica et Logopaedica*, 73(4), 316-325.
- Robbins, L. & Hoke, M. (2008). Using objective structured clinical examinations to meet clinical competence evaluation challenges with distance education students. *Perspectives in Psychiatric Care*. 44(2), 81-88.
- Schuwirth, L. & van der Vleuten, C. (2012). Programmatic assessment and Kane's validity perspective. *Medical Education*, 46(1), 38–48.
- Sheepway, L., Lincoln, M. & Togher, L. (2011). An international study of clinical education practices in speech-language pathology. *International Journal of Speech-Language Pathology*, 13(2), 174–185.
- Shehata, M., Kumar, A., Arekat, M., Alsenbesy, M., Al Ansari, A., Atwa, H., Ahmed, S., and Deifalla, A. (2020). A toolbox for conducting an online OSCE. *The Clinical Teacher*, 18(3), 236-242.
- Stern, D. (1996). Values on call: A method for assessing the teaching of professionalism. *Academic Medicine*, 71(10), S37–S39.
- St-Onge, C., Young, M., Eva, K., & Hodges, B. (2017). Validity: one word with a plurality of meanings. *Advances in Health Sciences Education*. 22(4), 853-867.
- Swanson, D., Clauser, B., & Case, S. (1999). Clinical skills assessment with standardized patients in high-stakes tests: a framework for thinking about score precision, equating, and security. *Advances in Health Science Education*, 4(1), 67–106.
- Tavares, W., Brydges, R., Myre, P., Prpic, J., Turner, L., Yelle, R. & Huiskamp, M. (2018). Applying Kane's validity framework to a simulation-based assessment of clinical competence. *Advances in Health Sciences Education*, 23(2), 323-338.