



## A validity perspective on interviews as a selection mechanism for entry to initial teacher education programmes

Paula Lehane, Zita Lysaght & Michael O'Leary

To cite this article: Paula Lehane, Zita Lysaght & Michael O'Leary (2021): A validity perspective on interviews as a selection mechanism for entry to initial teacher education programmes, European Journal of Teacher Education, DOI: [10.1080/02619768.2021.1920920](https://doi.org/10.1080/02619768.2021.1920920)

To link to this article: <https://doi.org/10.1080/02619768.2021.1920920>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 28 Apr 2021.



[Submit your article to this journal](#)



Article views: 1949



[View related articles](#)



[View Crossmark data](#)

# A validity perspective on interviews as a selection mechanism for entry to initial teacher education programmes

Paula Lehane <sup>a</sup>, Zita Lysaght<sup>b</sup> and Michael O'Leary<sup>a</sup>

<sup>a</sup>Centre for Assessment Research, Policy and Practice in Education (CARPE), Dublin City University, Dublin, Ireland; <sup>b</sup>Institute of Education, School of Policy and Practice, Dublin City University, Dublin, Ireland

## ABSTRACT

Across the world, teacher quality has come to be recognised as one of the most important variables affecting student outcomes; consequently, the regulation of entry into the profession is the subject of iterative review. The traditional 'one-off' interview, involving an interviewee and two or more interviewers, is a common, but not unproblematic, selection mechanism in the field. In particular, the modest positive correlation between performance at interviews and in clinical settings raises questions about using interviews as a selection mechanism for Initial Teacher Education (ITE) programmes. In this paper, we draw on validity theory and some key commentaries and studies in the research literature to offer a perspective on the extent to which the traditional interview provides data that can be used to make good decisions about applicants for ITE. The paper proposes a validity-based framework for use by practitioners to enhance the conceptualisation, design and evaluation of interviews in the process of teacher selection.

## ARTICLE HISTORY

Received 5 June 2020  
Accepted 17 April 2021

## KEYWORDS

Teacher selection; interviews; teacher education policy; Initial Teacher Education (ITE)

## Introduction

Educational research has consistently shown that teachers exert considerable effect on students' academic progress as well as their social and emotional well-being (e.g. Slater, Davies, and Burgess 2009). Even when controls are put in place to account for factors such as student background and prior attainment, having a 'good' teacher is one of the most important predictors of student success (Slater, Davies, and Burgess 2009). Consequently, the process of selecting the most suitable candidates for Initial Teacher Education (ITE) programmes routinely features in educational discourse particularly in countries like Ireland, Finland and Scotland where the teaching profession carries high social prestige (European Commission 2013; O'Doherty and Harford 2018) and where demand for places on ITE programmes from high calibre applicants often exceeds supply.<sup>1</sup>

The role of selection methods in such contexts is to assess candidates' background variables (e.g. relevant experience), as well as their cognitive and non-cognitive attributes, to determine their overall suitability for the ITE programme and the profession itself (Klassen et al. 2020). It is widely accepted that while cognitive attributes can be evaluated

**CONTACT** Paula Lehane  paula.lehane2@mail.dcu.ie  Centre for Assessment Research, Policy and Practice in education (CARPE), Dublin City University, St. Patrick's Campus, Drumcondra, Dublin 9. D09DY00

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

in a reasonably straightforward manner using academic transcripts as proxies, identifying appropriate selection tools to evaluate non-cognitive attributes (e.g. motivation, confidence and personality) is more challenging (Klassen et al. 2020). In the context of determining the eligibility of ITE candidates to both concurrent and consecutive<sup>2</sup> ITE programmes, minimum academic entry requirements are often supplemented by other mechanisms that are assumed to evaluate candidates' non-cognitive attributes (see Darmody and Smyth 2016). In Finland, for example, the Primary School Teacher Education programme involves a two-stage process that begins with an initial screening of applicants' academic study skills followed by an aptitude test usually in the form of an interview in which teacher educators meet face-to-face with applicants (Mankki, Mäkinen, and Pekka 2020). Darmody and Smyth (2016, 124) note that such interviews, which are also employed in Ireland and Scotland, are highly valued as mechanisms in many jurisdictions to gain insight into candidates' non-cognitive attributes such as personality and communication skills.

However, recent research on the efficacy of interviews as a selection measure for ITE programmes (e.g. Holden and Kitchen 2017; Klassen and Kim 2019) raises questions about the validity of inferences based on such measures. In response, we begin the next section of this paper by revisiting the interview as a measurement tool in order to highlight some of the challenges specific to ITE admission that the tool presents. In considering ways to improve such interviews, we consider, briefly, a number of developing research-based alternatives before proposing a design framework, informed by a construct validity perspective, to support practitioners in crafting, administering and interpreting interviews in the context of ITE.

## Interviews as a selection mechanism

Goho and Blackman (2006, 336) describe the interview as a 'form of measurement instrument' designed to predict success, evaluate 'non-cognitive skills' (e.g. interpersonal skills, communication skills), clarify the contents of a written application and provide a 'public relations' service. A 'traditional interview', as described by the authors, involves at least one (but usually two or more) interviewers asking a series of questions of the interviewee. Within the employment field, these interviews range from loosely structured (minimal guidelines), to moderately structured (predetermined scoring and question outlines) to highly structured (predefined questions and sample answers, training for interviewers) (Macan 2009; Goho and Blackman 2006). The key feature of the interview process is that it involves interactions between people in a social context. This means that both applicant and interviewer characteristics, such as personality, gender, race, age, disability, physical appearance, anxiety levels and pregnancy status, influence the value of an interview as a selection tool (Derous et al. 2016; Macan 2009). These characteristics have been associated with 'fast and frugal judgements ... during the initial impression formation process' (Derous et al. 2016, 105) and potential interview bias. As elaborated by Goho and Blackman (2006, 336):

... a number of interviewer tendencies could lead to bias, such as the halo effect, where the tendency is to generalize from one area to other areas, the impact of first impression being based primarily on appearance, the effect of the similarity or dissimilarity of interviewees to

interviewers, the tendency to place more weight on unfavourable information than favourable, and the rating bias of different interviewers with some exhibiting leniency, others severity and still others a central tendency bias. In addition, the interview process itself could have bias effects, including: the sequencing effect, that is, whether or not a candidate is preceded by several highly or lowly rated other candidates and the impact of the availability of varying pre-interview information on candidates.

While concern about bias features prominently in the field of organisational psychology (see Macan 2009, for example), researchers such as Goho and Blackman (2006, 336) have highlighted other relevant concerns when applying interviews to other areas, reminding us that ‘students are not employees’. Acknowledging the modest positive correlation between performance in interview and clinical settings, they stress the complex and diverse range of factors that can affect academic performance which, in turn, raises questions about interviews as selection mechanism for academic programmes such as consecutive programmes of ITE. Arguably, however, the most significant issue regarding the use of interviews to regulate entry to ITE relates to the lack of consensus within the field regarding *what* should be assessed or measured in interviews and *why* (Klassen et al. 2020). This raises the fundamental issue of construct validity.

## A validity perspective

According to the *Standards for Educational and Psychological Testing* (AERA, APA, NCME 2014, 11) ‘validity refers to the degree to which evidence and theory support the interpretations of test scores for the proposed uses of tests’. In this sense, a measurement instrument such as an interview cannot be considered valid or invalid. Rather, it is more appropriate to refer to the validity of a particular *interpretation* of data coming from a measurement instrument. Further, Newton (2017, 16) notes that the validity of an interpretation is dependent on the appropriateness of the procedures employed to generate the evidence used for interpretation. In essence, valid interpretations can only be drawn from appropriate evidence and, in respect to interviews, *construct validity evidence* is key.

Construct validity evidence is focused on the extent to which a tool, such as an interview, measures the ‘right’ psychological construct(s) (Messick 1989). As argued by Messick (1989), construct validity pertains to the extent to which a test is measuring the ‘right’ psychological construct (conscientiousness, for example). The case that a test, tool or measure has high construct validity is often made using evidence that these instruments use questions that relate to and represent some domain and that outcomes from them are systematically related to another measure (i.e. content-related and criterion-related validity evidence). In the specific case of interviews, this implies a need for clarity regarding the knowledge, skills or attributes that are being measured and the extent to which there is a correlation between performance at interview and some other measure or future performance. In light of a review by Hamdani, Valcea and Buckley (2014, 12), we also need to be mindful that ‘theoretical grounding is often times lacking in interview research’. As explained:

Previous validation efforts have generally relied on meta-analyses ... to clarify what constructs are measured by employment interviews. This approach runs the risk of confounding trait and method variances. Each method employed for testing a specific construct is a trait-

method unit. . . . Thus, different types of interview questions or interview methods in general will elicit responses that capture the intended constructs (e.g., proactive behaviors) as well as method-related constructs (e.g., impression management, in the case of certain types of interview questions). An interview score contaminated with method variance is less valid . . . .

Hamdani et al. (2014, 13) also make the important point that ‘meta-analyses cannot differentiate between intended construct-based variance and inherent method-based variance in the validity estimates’. The authors assert that differentiation between trait and method-based variance requires conceptual clarity at the design stage and an *a priori* knowledge of what is being measured. Bearing all this in mind, two questions arise now: What attributes of candidates are being measured at interviews and how well do interviews predict future performance (during the ITE programme, in the profession, or both)?

With regard to the first question, the use of interviews in ITE is based on the assumption that they are a suitable vehicle for assessing a candidate’s personal characteristics, specifically the ‘non-cognitive attributes’ needed to complete the programme of study and become effective teachers in the future (Klassen et al. 2020). Yet the range of non-cognitive skills targeted by ITE programmes at the point of selection using interviews is very broad. A review of the literature suggests that these non-cognitive attributes include confidence, integrity, resilience, motivation, empathy, organisation, and adaptability (Davies et al. 2016; Klassen et al. 2020). Investigating the popularity of interviews to assess wide ranging non-cognitive attributes, Davies et al. (2016) consulted key personnel involved in university-led ITE programmes in four different UK jurisdictions. Interesting, this research did not yield any evidence-based insights from participants regarding their decisions to employ interviews to determine candidates’ relevant skills and dispositions or, indeed, how their interview schedules assessed what they were purported to measure.

Of relevance to the second question regarding the predictive validity of ITE interviews for future professional performance is Klassen and Kim’s (2020, 33) contention that the selection process for ITE addresses the fundamental question of whether or not ‘it is likely that an [sic] applicant will be (or will become) an effective teacher’. Based on this perspective, the success of any selection process is based on the degree to which scores from a selection measure (e.g. interview) are associated with ‘true’ differences in teacher effectiveness as measured by an appropriate outcome measure (e.g. grade on teaching practicum). However, there are distinct construct validity issues at play here. First, teachers improve as they become more experienced and progress through ITE (e.g. Atteberry, Loeb, and Wyckoff 2015); this confounds the process of gathering validity evidence. Second, teacher effectiveness measures – irrespective of which and how many we use – are an *imperfect* representation of teacher effectiveness; disagreement persists regarding how this construct should be defined and measured. Measures used to capture the construct of teacher effectiveness include, among others, performance on observations, value-added measures and end of ITE grade-point averages (e.g. Klassen and Kim 2019; Goldhaber, Grout, and Huntington-Klein 2014). The variety of measures employed, coupled with disagreements over operational definitions, means that it can be difficult to associate admission interviews with teacher effectiveness. For example, Malvern (1991) compared the interview ratings of 112 preservice second-level teachers with their performance on teaching placement and coursework. While Malvern (1991) found a clear relationship between interview ratings and student teaching performance, no

relationship between interview rating and general course performance was found. This finding should be interpreted with caution as sample sizes for this study were relatively small. A larger Irish study ( $n > 330$ ) by Greaney, Burke and MacCann (1999) found that ratings on pre-entry interviews were significantly correlated with school placements grades. However, the effect size for this finding was relatively modest ( $r = .18$ ) and some concerns regarding the influence of extraneous variables (particularly the role of the ITE programme itself) within this study should be acknowledged.

From a validity perspective, then, it appears that there is a lack of clarity regarding both what a selection interview for ITE assesses and whether or not performance can be systematically related to the concept of teacher effectiveness. That said, it could be argued that this applies to all the selection tools in ITE, not just interviews. In a meta-analysis of 'the relation between selection practices at the point of hiring and subsequent teaching effectiveness' ( $n = 23$ ) and the 'the relation between selection into teacher education and teaching effectiveness during the ITE program' ( $n = 9$ ) conducted by Klassen and Kim (2019, 42), three categories of studies emerged based on the selection methods and tools employed: academic (e.g. undergraduate or secondary school grade-point average), non-academic (e.g. interview performance, personality measures) and both. Academic tools were assumed to assess cognitive skills and attributes while non-academic tools were assumed to assess a broad range of non-cognitive skills and attributes like interpersonal skills, suitability to the profession attitudes, behaviours, resilience etc. Outcome measures represented attempts to capture the construct of teacher effectiveness and included performance on professional and placement observations, student reading achievement, and end of ITE grade-point average.

Klassen and Kim's (2019) research confirms that current selection methods assessing cognitive and non-cognitive skills and constructs are statistically associated with common proxies for teacher effectiveness measures. The associations with the outcomes were significant for both academic ( $r = 0.14$ ) and non-academic selection tools ( $r = 0.10$ ) although the effect sizes were quite small, particularly when compared with the effect sizes obtained in similar reviews in other disciplines such as medicine. Further, the effect size for non-academic predictors ( $r = 0.10$ ) was significantly smaller than for academic predictors ( $r = 0.14$ ), suggesting that the challenge of reliably and validly evaluating non-cognitive attributes in high-stakes settings is hard to achieve using currently available selection tools.

From a construct validity perspective, this work suggests that current cognitive and non-cognitive based selection methods, including interviews, have good *potential* to select candidates for ITE but that work is needed to improve construct validity.

### **Efforts to improve the construct validity evidence of interviews**

A recurrent theme in the literature over the past two decades is the need for a more structured approach to interviews. More robust construct validity evidence for selection interviews is required and one way of achieving this is to clarify what is being assessed during the interview in a manner that minimises confounders such as interview bias. Three approaches of merit are considered here: the first focuses on recommendations for good practice, the second on structured interview schedules and the third on multiple mini interviews.

**Table 1.** Good practice guidelines for interviews.

Content-Based and Scoring Guidelines	Procedural Guidelines
Questions should be based on a job analysis	Extensive training for interviewers
Range of question types	Detailed notes should be taken
Longer interviews	Identical questions for all candidates
Limited number of follow up questions	Multiple interviewers should be involved
Detailed anchored rating scales	Same interviewers for all candidates
Each answer should be rated separately	Identical ancillary information on each candidate (e.g. pro-forma CV)
Statistical procedures should be used rather than interviewer judgements	Questions from candidates should come at the end of the interview

In 1997, Campion and colleagues published a much cited article identifying what they termed ‘Good Practice Guidelines’ for structured job interviews. As shown in Table 1, they made recommendations regarding the content of questions and how answers might be scored (column 1) allied with practical steps to be taken to ensure procedural integrity (column 2).

When protocols of this kind have been applied to ITE, research suggests they represent an improvement on the traditional, less structured alternative. A meta-analysis (Metzger and Wu 2008), for example, exploring the validity of the highly-structured *Teacher Perceiver Interview* (TPI) to inform hiring decisions in the US, reported a statistically significant relationship between TPI scores and indicators of teaching effectiveness (range of  $-0.12$  to  $0.87$ ) with a modest effect size (where  $N = 24$ ). Use of the TPI as a selection mechanism for ITE was not included in this study nor were comparable studies found at the time of writing this paper.

Informed by research and good practice guidelines (such as those recommended by Campion, Palmer, and Campion 1997), the medical profession has developed Multiple Mini-interviews (MMIs); highly structured interview protocols that may be applicable in an ITE setting. Designed to ‘dilute the impact of individual examiners and allow for a more generalizable aggregate performance rating’ (2004, 5), the MMI protocol consists of between six and 10 short interviews, each of seven to 10-minutes duration, conducted at different stations. Each station is staffed by at least one trained examiner who assesses a candidate’s skill level and overall suitability to the field of medicine using a 7-point Likert-type scale, ranging from ‘unsatisfactory’ to ‘outstanding’. Standardised questions or role-play scenarios are employed to tap into candidates’ non-cognitive skills such as conflict resolution and resilience. Subsequent evaluation research by Eva et al. (2004) and Jerant et al. (2012; 2018)) confirm the predictive value of the MMI over and above comparable alternatives such as the Objective Structured Clinical Examinations (OSCE), traditional interviews and grade point averages. Furthermore, a systematic literature of 41 studies conducted by Rees et al. (2016) examining MMI use for selection to a range of health programmes reported ‘reasonable’ feasibility, acceptability, validity, and reliability. Although the authors did not find evidence of significant bias against candidates on the basis of age, gender or socio-economic status when MMIs were conducted, Henderson et al. (2018) raised concerns about this issue. Reflecting on these findings, it can be suggested that there may be some merit in considering the use of modified MMIs in the context of ITE. However, the transferability of MMIs from a medical to a social science context may present difficulties.



## Applying a construct validity design framework for selection interviews for ITE

That valid screening procedures and selection tools need to be operationalised to manage entry of candidates into ITE is axiomatic. The literature reviewed in this paper has highlighted related difficulties with existing ITE interview selection mechanisms: an underlying lack of clarity about what is being assessed that undermines the validity of subsequent inferences and judgements. However, it has also identified some guidelines on how this issue might be progressed. From this, a design framework that foregrounds construct validity is now proposed to support practitioners in their use of interviews as a selection mechanism for ITE. As presented and discussed, the framework consists of a three step process relating to construct definition, instrument design and a validation strategy, respectively (Figure 1). At its heart is an assumption that the validity issues pertinent to the development of interviews for teacher selection outlined in this paper need to be addressed explicitly.

### Step 1: Define Constructs

As acknowledged by Darmody and Smyth (2016), screening and selection tools for ITE can only be effective if there is a clear understanding of the qualities and dispositions required for the profession. Work by Hamdani, Valcea, and Buckley (2014) claims that this is best achieved by first conducting a job task analysis and then explicating the theoretical links; their advice, on the need to distinguish between trait and method variances is also worth bearing in mind. A job task analysis (JTA) clarifies the real-world content of a job, as well as the requirements necessary for those who perform the job (Wolfe et al. 1991). A JTA is commonly used to create job descriptions that detail the tasks, knowledge, and skills required to fulfil a role to a minimally competent standard. It is recommended that this is done by relevant subject matter experts. The extant literature should then be used to

### A construct validity approach to interview design

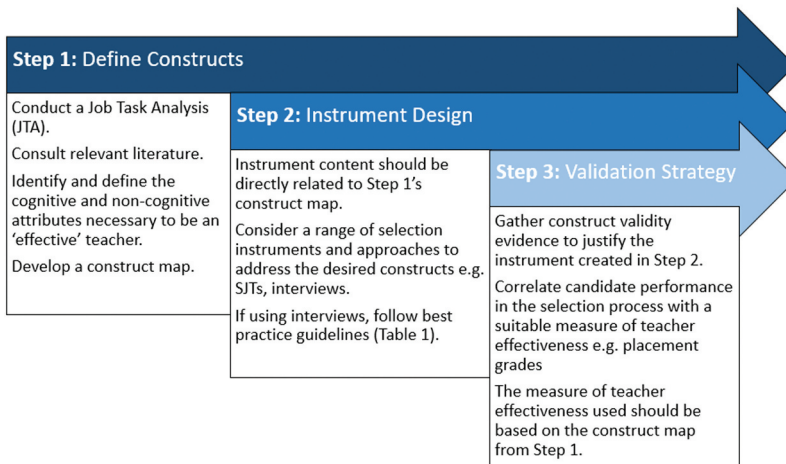


Figure 1. A construct validity approach to interview design.



supplement or confirm the statements arising from the JTA (Hamdani, Valcea, and Buckley 2014).

Given that selection measures aim to identify the underlying factors that could contribute to an applicant's development as an effective teacher in the future, clarity on the construct of 'teacher effectiveness' should be prioritised within the field. As noted by Klassen and Kim (2019, 34) 'teacher effectiveness refers to a set of within-person attributes – personality, motivation, beliefs, and dispositions – that interact with contextual factors (cultural, social, educational) to influence student outcomes'. Kunter et al.'s (2013) interactionist model attempts to clarify what attributes are needed to be a competent teacher. The model acknowledges that teaching competence develops over time (e.g. during ITE) and that within-person factors (e.g. academic attributes, non-academic attributes, background factors) can influence how an individual engages with the learning opportunities provided. Klassen et al. (2017) also advocate for a dynamic interactionist view to better understand the formation and development of effective teachers. In particular, research to identify the within-person factors that should be present at the point of entry to ITE to maximise future teaching effectiveness is ongoing. For example, work by Robertson-Kraft et al. (2014) found that a disposition to pursue challenging goals with sustained passion and perseverance is a personality factor that can significantly predict effectiveness (in terms of students' academic gains over a one-year period) and retention among novice teachers in low-income districts ( $d = .42$ ).

While methodological issues exist in research on teacher effectiveness (e.g. how teacher effectiveness should be defined), a clearer understanding of what personal attributes should be identified in interviews should enhance the validity of the evidence arising from this selection tool. For example, if interviews are to be used as a selection tool, the relevant authorities should explicitly state what this tool intends to screen candidates. This 'construct map' can then be used to design an effective interview schedule.

## ***Step 2. Instrument Design***

Given the limited evidence available on the overall reliability and validity of interviews as a selection tool within ITE, the efficacy of this approach to identify effective teachers is difficult to determine (e.g. Klassen and Kim 2019). Therefore, it could be argued that interviews should not be used as a selection tool for admission to ITE. However, there is no doubt that among academic personnel interviews remain a popular approach to selecting candidates for jobs such as teaching or for educational programmes such as master's degrees (Davies et al. 2016). Indeed, there are many who claim that applicants not only expect interviews to be part of a successful job search, but view them as being fairer than other selection procedures such as psychometric tests (Macan 2009). As a result, the removal of interviews from the screening process may be met with some resistance. If relevant stakeholders decide to persist with this selection tool, efforts should be made to improve them by following certain guidelines like those previously outlined by Campion, Palmer, and Campion (1997). The use of structured interview schedules should also be adhered to ensuring that the interview process is standardised and that the questions asked are related to the constructs being assessed. Alternatively, interview approaches such as MMIs may need to be considered. MMIs would allow an examination of a range of

constructs in a reliable manner that minimises the interviewer bias. Research conducted as part of the *Teacher Selection Project* (TSP 2019) as well as the experiences of bodies such as the Royal College of Surgeons in Ireland (RCSI; 2018) which has begun to use MMIs to determine medical students' suitability to different specialities may prove to be particularly useful given the relatively modest body of literature in the area currently.

If a construct map is available, this may allow ITE providers to decide which constructs to measure using interviews and which constructs to measure using alternative selection methods (Hamdani, Valcea, and Buckley 2014). For example, communication skills, if identified as a key construct for prospective teachers, may be best measured through a selection interview. However, other skills or attributes like conflict resolution skills may be more easily, and validly, measured through other selection tools. For example, there is a growing body of research (e.g. Klassen et al. 2020) to suggest that many of the non-cognitive skills and attributes that teaching institutions claim to measure using interviews may be better measured using situational judgment tests (SJTs). SJTs simulate realistic job-related situations where candidates are presented with task-based scenarios (in the form of text, images or videos) and asked to choose a response from a list of different options (McDaniel and Nguyen 2001). Faced with a complex scenario, a candidate's selected response is assumed to reveal information about some implicit non-cognitive attribute of the candidate, for example their level of empathy or adaptability (TSP 2019). This contrasts with how such traits are assessed in interviews where candidates are asked to self-report on their non-cognitive attributes. SJTs have been applied as a selection tool in a range of other sectors and have been associated with high levels of predictive and incremental validity and fewer in-group differences (Patterson et al. 2016).

Taking into consideration the inter-disciplinary evidence regarding the predictive validity of SJTs as a measure of non-cognitive ability and their suitability for use within the teaching profession, the TSP (2019) has begun the process of developing SJTs specifically for use in selection decisions for ITE. Klassen and Rushby (2019) administered traditional individual interviews, group tasks, maths and literacy tests and situational judgement tests (SJTs) to candidates as part of an ITE providers' selection process. A total of 132 applicants (Primary Teacher Candidates: 96; Second-Level Teacher Candidates: 36) participated. The results showed that the SJTs developed for primary and second-level ITE applicants were significantly predictive of preservice teachers' grades for their final teaching placement (Primary:  $r = .30$ ; Secondary:  $r = .35$ ). The other selection methods used at interview showed no such evidence of predictive validity. These findings may be interpreted as suggesting that SJTs are a highly effective ITE selection tool. Interestingly, Klassen and Rushby (2019) found that SJT scores used for entry to ITE were more predictive of final, rather than initial, teaching placements. This suggests that the predictive validity of SJTs increases over time, which is consistent with findings from research in other disciplines like dentistry (Buyse and Lievens 2011) and supports Kunter et al.'s (2013) interactionist model to explain teacher effectiveness.

### **Step 3. Validation Strategy**

If institutions wish to claim that a particular tool or instrument supports them in selecting appropriate candidates of ITE, the selection tool in question must be properly validated. If the interview is designed to identify specific attributes necessary for effective teaching

(e.g. motivation, as outlined in Step 1), sufficient construct validity evidence must be gathered. This would involve sampling future teaching performance to determine if the attributes measured in the interview were related to teaching efficacy. Research critically evaluating the effectiveness of different teacher selection tools, including interviews, is not common and what is available has been described as 'ad hoc' (Goldhaber, Grout, and Huntington-Klein 2014, 2) and 'information poor' (Liu and Johnson 2006, 324). Therefore, to help inform any decisions regarding the deployment of available teacher selection tools for entry to ITE, further evidence and information is essential.

Some work has been done to better understand the predictive validity of selection measures for teachers. However, the vast majority of this research, with the exception of Klassen and Kim's (2019) recent meta-analysis, has been centred around teacher employment rather than entry to ITE. For example, Goldhaber, Grout, and Huntington-Klein (2014) examined selection practices for teaching jobs in one district in the US, where teachers' applications contained background information (e.g. academic qualifications) and recommendation letters. The applications were evaluated using carefully constructed 21-point (general) and 60-point screening (job-specific) rubrics. The second of these rubrics was carefully constructed to reflect key constructs that would advance student outcomes (e.g. classroom management, instructional skills, flexibility). After this screening process, interviews were conducted by schools to identify who would be offered a job. They found that teachers with higher rubric scores also had higher scores on school-based evaluations/observations and their students showed improvements after one year in their performance on standardised tests for mathematics (but not English). While the authors acknowledged that other variables which we not included in the rubrics can also contribute to teacher effectiveness, their work remains a good example of construct validity evidence gathered to justify the use of a 'double' screening process in a particular district in the US. Interestingly, Goldhaber, Grout, and Huntington-Klein (2014) did not examine if the interview stage was helpful in the selection process as there were no set interview evaluation criteria and the principal had discretion over the content and structure of the interview.

While there is some evidence to justify the use of certain selection tools for teachers, limited evidence has been gathered to support the use of interviews. This should be rectified to ensure that institutions can be confident in their selection decisions about entry to ITE. This underscores the need for educational institutes that offer ITE programmes to ensure that their interviews are appropriately conceptualised (Step 1) and designed (Step 2). Assuming this requirement is met, attempts to determine the efficacy of interview performance by correlating candidates' scores with some measure of teacher effectiveness is recommended. For ITE, performance on school placements could be used as a proxy measure for teaching effectiveness. However, the process of assigning grades on school placement would have to be carefully examined to ensure that it aligns with the constructs considered in Step 1.

One small-scale study by Ebmeier and Ng (2006) appears to have made some effort to engage in this three-step process to ensure the construct validity of interviews as a selection tool for hiring teachers. Ebmeier and Ng (2006) developed a computer-adaptive interview instrument that aimed to provide administrators in urban school districts in the US with an improved interview instruments for identifying effective teaching candidates for their particular setting. Key themes that were important for

success in the urban classroom in the US were identified from the extant literature and in consultation with subject matter experts. The interview schedule was designed with these key constructs in mind (e.g. knowledge of urban schools/communities, interactions with urban students). Rubrics were designed to assist in scoring interviewees' responses. By comparing the phone-interview scores of 30 teachers with varying effectiveness ratings provided by administrators in one urban district, significant correlations were found. Regression analysis indicated a significant amount of variance in teachers' effectiveness ratings (28%) could be predicted from their scores on the interview instrument. Although there were some limitations with this study that jeopardise the reliability of its results (e.g. use of phone interviews, limited definition and measure of teacher effectiveness), it does demonstrate how a three-step approach to gathering construct validity evidence can increase confidence in the selection decisions made at interview.

## Conclusion

This paper proposes a validity-based framework for use by education practitioners to enhance the manner in which they conceptualise, design and evaluate interviews in the process of teacher selection. The intention of this framework is to bolster confidence in the use of information from interviews when assessing the suitability of candidates for teaching as a career, in the immediate and long-term. In order to judge the value of the proposed framework, a programme of research will be needed. It is recommended that, as a starting point, this might involve colleagues across institutions nationally and/or internationally who currently use interviews as part of their screening mechanisms. Unfortunately, interest in conducting such research may be limited to those countries who already use interviews as a selection mechanism. Yet, it is hoped that the contents of this article can still contribute to the general theoretical discussion of teacher selection practices and how validity should be at the core of such debates.

For those countries that do use interviews as a selection tool and assuming the continued over-subscription of candidates to ITE in those jurisdictions (e.g. Ireland, Finland, Scotland), the validity and reliability of evidence used to inform candidate selection remains a priority. It is suggested that these ITE providers give consideration, in the immediate term, to the range of interview strategies discussed in this paper which may enhance the validity and reliability of the decisions they make. Developments in the admission protocols used in the medical sciences, and the use of MMI's in particular, merit review. Although, as acknowledged previously, the transferability of practices from a medical to an education context cannot be assumed and, indeed, may prove financially prohibitive.

In line with the framework outlined in this paper, the first step – and the over-arching research priority within the field itself – must be to clarify the core construct(s) being assessed when selecting candidates for ITE. While some efforts have been made to achieve this (e.g. Kunter et al. 2013), consensus on what constructs underlie effective teaching has yet to be reached. Meta-studies analysing previous research into the relationship between teacher effectiveness measures and student outcomes can be a starting point in addressing this issue. Once the constructs underlying teacher effectiveness have been mapped, researchers and practitioners can then turn their attention to the design of selection tools for the field of ITE using findings from peer-reviewed research as a guide. While the interview may be the

obvious tool to 're-design' given its long history of use, the potential of other approaches such as SJTs cannot be ignored in the short term. Of particular importance will be the need to conduct validation studies involving all new or modified approaches in different national and international contexts. The goal of such a research agenda must be to provide robust validity evidence to inform our understanding of the important task of selecting the best candidates for ITE programmes and, ultimately, the teaching profession itself.

## Notes

1. For example, 252 eligible candidates were invited to interview for 60 places that were available on the Professional Masters in Education (primary) programme at one Irish provider of ITE in 2019 (ML, personal communication, December 2019).
2. Concurrent programmes of ITE combine an undergraduate and professional education degree into one extended programme whereas a consecutive programme is open for candidates who already hold an undergraduate degree (Darmody and Smyth 2016).

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

The funding text should read: This work was partially supported by a grant from the Conference of Heads of Irish Colleges of Education (CHOICE).

## Notes on contributors

*Paula Lehane* is a PhD candidate in Dublin City University (DCU) at the Centre for Assessment Research, Policy and Practice in Education (CARPE). She previously worked as a primary school teacher and was the Special Educational Needs (SEN) coordinator of a large urban primary school. Her research interests include SEN, educational assessment and technology-based assessments. <https://orcid.org/0000-0003-0856-3505>

*Zita Lysaght* is a member of the School of Policy and Practice and a Research Associate and member of the Advisory Board and Advisory Panel of the Centre for Assessment Research, Policy and Practice (CARPE) at the Institute of Education, DCU. She coordinates and teaches classroom assessment and research methodology modules on undergraduate, masters and doctoral programmes and directs and supervises a range of research and doctoral projects.

*Michael O'Leary* holds the Prometric Chair in Assessment at DCU and is Director of the Centre for Assessment Research, Policy and Practice (CARPE) at the Institute of Education, DCU. He leads a programme of research at CARPE focused on assessment in education and in the workplace. <https://orcid.org/0000-0002-6771-904X>

## ORCID

Paula Lehane  <http://orcid.org/0000-0003-0856-3505>

## References

- American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME). 2014. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Atteberry, A., S. Loeb, and J. Wyckoff. 2015. "Do First Impressions Matter? Predicting Early Career Teacher Effectiveness." *AERA Open* 1: 233285841560783. (October 2015). doi: [10.1177/2332858415607834](https://doi.org/10.1177/2332858415607834).
- Buysse, T., and F. Lievens. 2011. "Situational Judgment Tests as a New Tool for Dental Student Selection." *Journal of Dental Education* 75 (6): 743–749. doi:[10.1002/j.0022-0337.2011.75.6.tb05101.x](https://doi.org/10.1002/j.0022-0337.2011.75.6.tb05101.x).
- Campion, M., D. Palmer, and J. Campion. 1997. "A Review of Structure in the Selection Interview." *Personnel Psychology* 50: 655–702. doi:[10.1111/j.1744-6570.1997.tb00709.x](https://doi.org/10.1111/j.1744-6570.1997.tb00709.x).
- Darmody, M., and E. Smyth. 2016. *Entry to Programmes of Initial Teacher Education*. Dublin: Economic and Social Research Institute. <https://www.teachingcouncil.ie/en/Publications/Teacher-Education/Documents/Entry-Requirements.pdf>
- Davies, P., M. Connolly, J. Nelson, M. Hulme, J. Kirkman, and C. Greenway. 2016. "Letting the Right One In: Provider Contexts for Recruitment to Initial Teacher Education in the United Kingdom." *Teaching and Teacher Education* 60: 291–302. doi:[10.1016/j.tate.2016.09.003](https://doi.org/10.1016/j.tate.2016.09.003).
- Derous, E., A. Buijsrogge, N. Roulin, and W. Duyck. 2016. "Why Your Stigma Isn't Hired: A Dual-process Framework of Interview Bias." *Human Resource Management Review* 26: 90–111. doi:[10.1016/j.hrmr.2015.09.006](https://doi.org/10.1016/j.hrmr.2015.09.006).
- Ebmeier, H., and J. Ng. 2006. "Development and Field Test of an Employment Selection Instrument for Teachers in Urban School Districts." *Journal of Personnel Evaluation in Education* 18 (3): 201–218. doi:[10.1007/s11092-006-9021-4](https://doi.org/10.1007/s11092-006-9021-4).
- European Commission. 2013. "Study on Policy Measures to Improve the Attractiveness of the Teaching Profession in Europe." Luxembourg: Publications Office of the European Union. [https://ec.europa.eu/assets/eac/education/library/study/2013/teaching-profession1\\_en.pdf](https://ec.europa.eu/assets/eac/education/library/study/2013/teaching-profession1_en.pdf)
- Eva, K., H. Reiter, J. Rosenfeld, and G. Norman. 2004. "The Ability of the Multiple Mini-interview to Predict Preclerkship Performance in Medical School." *Academic Medicine* 79 (10): 540–542. doi:[10.1097/00001888-200410001-00012](https://doi.org/10.1097/00001888-200410001-00012).
- Goho, J., and A. Blackman. 2006. "The Effectiveness of Academic Admission Interviews: An Exploratory Meta-analysis." *Medical Teacher* 28 (4): 335–340. doi:[10.1080/01421590600603418](https://doi.org/10.1080/01421590600603418).
- Goldhaber, D., C. Grout, and N. Huntington-Klein. 2014. *Screen Twice, Cut Once: Assessing the Predictive Validity of Teacher Selection Tools*. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research. <https://files.eric.ed.gov/fulltext/ED560673.pdf>
- Greaney, V., A. Burke, and J. McCann. 1999. "Predictors of Performance in Primary School Teaching." *The Irish Journal of Education* 30 (2): 22–37. <http://www.erc.ie/documents/vol30chp2.pdf>
- Hamdani, M. R., S. Valcea, and M. R. Buckley. 2014. "The Relentless Pursuit of Construct Validity in the Design of Employment Interviews" *Human Resource Management Review* 24 (2): 160–176. doi:[10.1016/j.hrmr.2013.07.002](https://doi.org/10.1016/j.hrmr.2013.07.002).
- Henderson, M., C. Kelly, E. Griffin, T. Hall, A. Jerant, E. Peterson, J. Rainwater, F. Sousa, D. Wofsy, and P. Franks. 2018. "Medical School Applicant Characteristics Associated with Performance in Multiple Mini-interviews versus Traditional Interviews: A Multi-institutional Study." *Academic Medicine* 93 (7): 1029–1034. doi:[10.1097/acm.0000000000002041](https://doi.org/10.1097/acm.0000000000002041).
- Holden, M., and J. Kitchen. 2017. "Choosing How We Choose: A Review of Teacher Education Admissions Literature." *Teaching and Learning* 11 (1): 16–31. doi:[10.26522/tl.v11i1.436](https://doi.org/10.26522/tl.v11i1.436).
- Jerant, A., E. Griffin, J. Rainwater, M. Henderson, F. Sousa, K. Bertakis, J. Fenton, and P. Franks. 2012. "Does Applicant Personality Influence Multiple Mini-interview Performance and Medical School Acceptance Offers?." *Academic Medicine* 87: 1250–1259. doi:[10.1097/ACM.0b013e31826102ad](https://doi.org/10.1097/ACM.0b013e31826102ad).
- Jerant, J., M. Henderson, E. Griffin, T. Hall, C. Kelly, E. Peterson, D. Wofsy, D. Tancredi, F. Sousa, and P. Franks. 2018. "Do Admissions Multiple Mini-interview and Traditional Interview Scores Predict



- Subsequent Academic Performance? A Study of Five California Medical Schools." *Academic Medicine* September: 3–28. doi:10.1097/ACM.0000000000002440.
- Klassen, R. M., and L. E. Kim. 2017. "Assessing Critical Attributes of Prospective Teachers: Implications for Selection into Initial Teacher Education Programs." In *British Journal of Educational Psychology Monograph Series II: Psychological Aspects of Education*, edited by D. W. Putwain and K. Smart, 5–22. Oxford: Wiley. [http://eprints.whiterose.ac.uk/124823/1/BJEP\\_Final\\_14.01.17.pdf](http://eprints.whiterose.ac.uk/124823/1/BJEP_Final_14.01.17.pdf)
- Klassen, R. M., and J. Rushby. 2019. "Can We Predict Preservice Teachers' Performance in Teaching Placements? The Validity of ITE Selection Methods." *Teacher Selection Project Working Paper*. [https://www.teacherselect.org/media/uploads/2019/01/18/tsp-working-paper-january-2019-can-we-predict-preservice-teachers-performance-in-teaching-placements\\_-19448.pdf](https://www.teacherselect.org/media/uploads/2019/01/18/tsp-working-paper-january-2019-can-we-predict-preservice-teachers-performance-in-teaching-placements_-19448.pdf)
- Klassen, R. M., and L. E. Kim. 2019. "Selecting Teachers and Prospective Teachers: A Meta-analysis." *Educational Research Review* 26: 32–51. doi:10.1016/j.edurev.2018.12.003.
- Klassen, R. M., L. E. Kim, J. V. Rushby, and L. Bardach. 2020. "Can We Improve How We Screen Applicants for Initial Teacher Education?." *Teaching and Teacher Education* 87: 1–11. doi:10.1016/j.tate.2019.102949.
- Kunter, M., T. Kleickmann, U. Klusmann and D. Richter. 2013. „The development of teachers' professional competence„. In *Cognitive activation in the mathematics classroom and professional competence of teachers* edited by Kunter, M., J. Baumert, W. Blum, U. Klusmann, S. Krauss, M. Neubrand, 63–77. New York: Springer
- Liu, E., and S. M. Johnson. 2006. "New Teachers' Experiences of Hiring: Late, Rushed, and Information Poor." *Educational Administration Quarterly* 42: 324–360. doi:10.1177/0013161X05282610.
- Macan, T. 2009. "The Employment Interview: A Review of Current Studies and Directions for Future Research." *Human Resource Management Review* 19: 203–218. doi:10.1016/j.hrmr.2009.03.006.
- Malvern, D. 1991. "Assessing the Personal Qualities of Applicants to Teacher Training." *Studies in Educational Evaluation* 17: 239–253. doi:10.1016/S0191-491X(05)80082-3.
- Mankki, V., M. Mäkinen, and R. Pekka. 2020. "Teacher Educators' Predictability and Student Selection Paradigms in Entrance Examinations for the Finnish Primary School Teacher Education Programme." *European Journal of Teacher Education* 43 (2): 151–164. doi:10.1080/02619768.2019.1672653.
- McDaniel, M. A., and N. T. Nguyen. 2001. "Situational Judgement Tests: A Review of Practice and Constructs Assessed." *International Journal of Selection and Assessment* 9 (1/2): 103–113. doi:10.1111/1468-2389.00167.
- Messick, S. 1989. "The once and Future Issues of Validity: Assessing the Meaning and Consequences of Measurement." In *Test Validity*, edited by H. Wainer and H. Braun, 33–45. New Jersey: Lawrence Erlbaum Associates.
- Metzger, S. A., and M.-J. Wu. 2008. "Commercial Teacher Selection Instruments: The Validity of Selecting Teachers through Beliefs, Attitudes, and Values." *Review of Educational Research* 78: 921–940. doi:10.3102/0034654308323035.
- Newton, P. E. (2017). "An Approach to Understanding Validation Arguments". *Report for the Office of Qualifications and Examinations Regulation* (Ofqual). [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/653070/An\\_approach\\_to\\_understanding\\_validation\\_arguments.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/653070/An_approach_to_understanding_validation_arguments.pdf)
- O'Doherty, T., and J. Harford. 2018. "Teacher Recruitment: Reflections from Ireland on the Current Crisis in Teacher Supply." *European Journal of Teacher Education* 41 (5): 654–669. doi:10.1080/02619768.2018.1532994.
- Patterson, F., A. Knight, J. Dowell, S. Nicholson, F. Cousans, and J. Cleland. 2016. "How Effective are Selection Methods in Medical Education? A Systematic Review." *Medical Education* 50 (1): 36–60. doi:10.1111/medu.12817.
- Rees, E., A. Hawarden, G. Dent, R. Hays, J. Bates, and A. Hassell. 2016. "Evidence regarding the Utility of Multiple Mini-interview (MMI) for Selection to Undergraduate Health Programs: A BEME Systematic Review." *Medical Teacher* 38 (5): 443–455. doi:10.3109/0142159X.2016.1158799.
- Robertson-Kraft, C., A. L. Duckworth, E. K. Penner, and A. Conley. 2014. "True Grit: Trait-level Perseverance and Passion for Long-term Goals Predicts Effectiveness and Retention among



Novice Teachers." *Teachers College Record* 116: 1–27. <http://www.tcrecord.org/Content.asp?ContentId=17352>

Royal College of Surgeons Ireland (RCSI). 2018. "Core Surgical Training: Trainee Guidebook." <https://msurgery.ie/wp-content/uploads/2018/06/Core-SurgicalTraining-GuideBook-V7-FINAL-June-2018.pdf>

Slater, H., N. Davies, and S. Burgess. 2009. "Do Teachers Matter? Measuring the Variation in Teacher Effectiveness in England." Centre for Market and Public Organisation: University of Bristol. [www.bristol.ac.uk/compo/publications/papers/2009/wp212.pdf](http://www.bristol.ac.uk/compo/publications/papers/2009/wp212.pdf)

Teacher Selection Project (TSP). (2019). "Teacher Selection Project." <https://www.teacherselect.org/>

Wolfe, P., M. Wetzel, G. Harris, T. Mazour, and J. Riplinger. 1991. *Job Task Analysis: Guide to Good Practice*. Educational Technology Publications: USA.