



*Teacher Selection Project Working Paper*  
18.01.19

## **Can We Predict Preservice Teachers' Performance in Teaching Placements? The Validity of ITE Selection Methods**

Robert M. Klassen and Jade V. Rushby  
University of York, UK

This *Working Paper* reports pre-print work that is not yet published (but is submitted for publication). The purpose of the *Working Paper* is to stimulate discussion and to contribute to the advancement of knowledge. Please check our [project website](#) for news of recent publications. This work is supported by the European Research Council (grant #647234 SELECTION)

## **Can We Predict Preservice Teachers' Performance in Teaching Placements? The Validity of ITE Selection Methods**

Robert M. Klassen & Jade V. Rushby

### **Abstract**

The purpose of this paper is to evaluate the relationship between Initial Teacher Education (ITE) program selection methods and candidates' subsequent performance (up to 18 months later) in two teaching placements in primary and secondary schools. Individual interviews, group tasks, math and literacy tests (for primary applicants only) and situational judgment tests (SJTs) were administered to candidates as part of an ITE providers' interview process. A total of 132 applicants (96 primary and 36 secondary) took part in the process. The results showed that the SJTs developed for primary and secondary ITE applicants were significantly predictive of preservice teachers' grades for the final teaching placement, but not for the introductory placement. Other selection methods used at interview (face-to-face interviews, group tasks, math and literacy tests) did not predict scores on either of the two teaching placements. Applicant reactions to the SJTs were generally favourable. The results have the potential to influence teacher selection practices.

### **Introduction**

One of the goals of the selection process for entry into initial teacher education (ITE) is to identify those applicants who are most likely to experience success during their training program. Success in an ITE program can be measured in multiple ways: retention in the program, academic performance on coursework, development of reflective awareness of teaching practices and of the profession, and perhaps most importantly, teaching competence displayed during teaching practice opportunities. In most ITE programs, successful completion of the program requires completion of one or more extended teaching placements (e.g., Cohen, Hoz, & Kaplan, 2013). For most preservice teachers, the ITE teaching placement is the most realistic approximation of their future work as teachers. A report by the National Council for Accreditation of Teacher Education (NCATE; 2010) suggests that student teaching is one of the most important aspects of an ITE program, and empirical evidence suggests that practice teaching experiences are one of the best predictors of success in actual teaching in the first professional position (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009). However, identifying those who will experience success in their preservice teaching placements is no easy task, with few studies critically evaluating how methods used to select applicants into the ITE program are related to successful completion of preservice teaching placements (Klassen & Kim, 2019).

The effectiveness of a teacher and the quality of the instruction they provide can have a substantial impact on the academic achievement and future prospects of their students (Chetty et al., 2010; Hanushek & Zhang, 2009; Hanushek & Rivkin, 2010; Murnane, Willett, Duhaldeborde, & Tyler, 2000; Schumaker, Grigsby, & Vesey, 2013). A central challenge facing many Initial Teacher Education (ITE) providers is how to select candidates during the interview process who have the potential to succeed in their training programs, and become effectual teachers who have a positive impact on student outcomes (Borman & Dowling, 2008; Rockoff, Jacob, Kane, & Staiger, 2011). As part of the selection process, candidates are frequently assessed on their cognitive abilities (such as, subject knowledge and academic skills), and non-cognitive attributes (such as, personality and motivation to teach); as evidence suggests these attributes can have a joint impact on future success in the field (Rockoff et al., 2011). Although cognitive attributes can be fairly readily assessed by ITE providers by looking at applicants' academic credentials (e.g. A-levels, undergraduate degree); non-cognitive attributes can be harder to assess and measure in the interview

process (Duckworth & Yeager, 2015). A recent analysis by Klassen & Kim (2019) shows that the correlation between non-cognitive attributes assessed at selection and subsequent teaching performance is not very strong ( $r = .10$ ).

Research in a variety of other sectors, such as medicine (Lievens, Patterson, Corstjens, Martin, & Nicholson, 2016; Patterson et al., 2016), dentistry (Buyse & Lievens, 2011), and business (Cook, 2016), has found that certain selection instruments targeting non-cognitive attributes, such as situational judgment tests (SJTs), may be more effective than traditional selection tools, such as interviews, in predicting applicants' future performance in the field. SJTs assess candidates' internal traits and behavioural tendencies by presenting candidates with context-rich scenarios and asking them to select what they think they should do in a given situation from a set number of options (Lievens, Peeters, & Schollaert, 2008). In the education sector, comparatively little research has been conducted in relation to the development and testing of evidence-based selection methods, such as SJTs (Bowles, Hattie, Dinham, Scull, & Clinton, 2014; Hanushek, 2014). The purpose of this article is to explore the relations between ITE program selection methods and preservice teachers' performance in two extended teaching placements. Results from the study have the potential to influence preservice teacher selection practices.

### **Initial Teacher Education Selection Methods**

ITE selection methods can dramatically differ between institutions in terms of their cost to run, ability to fairly select applicants, and to effectively predict applicants' performance outcomes on the program and in professional settings (Hobson, Ashby, McIntyre, Malderez, 2010; Lievens, Peeters, & Schollaert, 2008; Casey & Childs, 2007). The majority of ITE providers select prospective teachers based upon applicants' academic credentials (e.g. A-levels, undergraduate degree), personal statements, and their performance in interviews and/ or group tasks (Casey & Childs, 2007; Lievens et al., 2008). This process can be time-consuming and expensive in terms of labour and resources, especially for ITE providers with high numbers of applicants (Casey & Childs, 2007; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Metzger & Wu, 2008). Selection methods, such as interviews, are frequently designed by the ITE provider and aim to assess attributes deemed to be important for effectual teaching practice. Interviews are one of the most frequently used methods in selecting candidates for most professions (Cook, 2016), however the validity and reliability of interviews has been found to vary depending on the format (for instance, structured vs. unstructured) and content (i.e. situational vs job related) (McDaniel et al., 1994; Byrnes et al., 2000; Denner et al., 2001). For example, McDaniel et al. (1994) conducted a meta-analysis of research on interviews, and ascertained that unstructured interviews had a lower validity in comparison to structured interviews, and that interviews that were situational were more valid in comparison to traditional job-related interviews.

Although structured interviews may be more predictive of future performance, they can still be time consuming to complete, expensive, and have limited predictive validity (Metzger & Wu, 2008; Cook, 2016). Metzger and Wu (2008) assessed the Teacher Perceiver Interview and its online format, TeacherInsight, which is a commercial interview method used by schools to select teachers in 15% of school districts in the US. Initial start-up costs, annual fees, and training were found to be expensive, and the interviews took between 40 mins (TeacherInsight) to 2 hours (Teacher Perceiver Interview) to conduct, and validity was found to be moderate (.28) (Metzger & Wu, 2008). Furthermore, interviews have frequently been found to be subjectable to interviewer bias (Davison & Burke, 2000) and may not accurately measure target attributes but other personal factors such as the likeability of the applicant (Schumacher et al., 2013). Thus, current ITE selection methods may be expensive to conduct and may not objectively and systematically measure non-cognitive qualities that may lead to future success in the teaching profession.

## Situational Judgment Tests (SJTs)

SJTs have been utilised in a range of other fields, and have been found to have good predictive validity in terms of candidates' future performance (Buyse & Lievens, 2011; Lievens et al., 2008; Patterson et al., 2016). SJTs aim to assess candidates' underlying values, behavioural tendencies, and procedural knowledge, and may complement existing teacher selection methods by providing an objective assessment of candidates' non-cognitive skills (Ahmed, Rhydderch, & Matthews, 2012). SJTs may also be cost and time effective in comparison to other methods such as assessment centres, as candidates can complete the test online with limited to no invigilation (Ahmed et al., 2012; Lievens et al., 2008). Additionally, the scoring key for SJTs is developed by Subject Matter Experts (SMEs), which ensures scoring is universal across candidates and not influenced by interviewers' assessments of candidates' performance (Lievens et al., 2008; Salvatori, 2001). In relation to medical education, Patterson et al. (2016) ascertained that situational judgment tests (SJTs) and assessment centres were more effective methods for assessing non-cognitive attributes in comparison to interviews, personal statements and references; as they showed higher predictive validity and were less likely to be influenced by interviewers' unconscious biases. Several studies have ascertained that the predictive validity of SJTs may also increase over time (Buyse & Lievens, 2011; Lievens & Sackett, 2012). For example, in relation to dentistry, Buyse and Lieven (2011) found that the predictive validity of SJTs significantly increased over the course of training, although the predictive validity of cognitive ability assessments decreased over the same period. In the education sector, limited longitudinal data has been collected to assess the reliability and validity of popular teacher selection instruments in terms of their ability to effectively predict ITE students' future performance (Goldhaber, Grout, Huntington-Klein, 2015; Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009). Therefore, there may be potential value in assessing the relationship between teacher selection processes and student outcomes in order to determine the effectiveness of current selection methods and to inform the future development of selection instruments.

## Teacher Education Placements and the Current Study

Practical teaching placements are an integral part of assessments in ITE programs, with students being appraised on their performance in at least two extended school placements on most ITE courses (Carter, 2015). Students may spend up to half of their course in a school placement, and are assessed on their teaching ability, reflectiveness, and lesson planning (Chung, 2008). A study conducted by Boyd et al. (2009) into aspects of ITE training that effectively prepared prospective teachers for teaching, and established that practical teaching practice was associated with students' level of effectiveness in their first year of teaching. This may suggest that students' performance in placements may be an indicator of their effectiveness as Newly Qualified Teachers. However, it can be challenging at the interview stage to identify candidates who will perform well in the practical element of the course, and limited consideration has been given to comparing selection processes with students' practical placement performance. Therefore, the central aim of this study is to address the following:

- 1) What is the relationship between ITE selection processes and applicants' subsequent performance on teaching placements?
- 2) Are SJT scores, collected as part of the ITE selection process, associated with students' placement grades?

## Method

### Participants

**Primary applicants.** There were 838 primary ITE applicants (mean age was 20.03 years, 80.31% were female, and 94.63% White British) completed the SJTs following their interviews at a higher education teacher training provider between November 2016 and July

2017. Of these candidates, 96 primary applicants, with a mean age of 23.2 years ( $SD=3.82$ ), were accepted and completed the Post Graduate Certificate of Education program. 97.9% of participants identified as being White British, 1% Asian or Asian British, and 1% mixed/multiple ethnic groups. 100% of participants reported having prior experience working in schools either in a voluntary (60.4%), or professional (4.2%) capacity, or both (35.4%). In terms of length of prior experience, 6.3% of participants had 1-4 weeks experience, 24.2% had 1-3 months, 19.8% had 3-6 months, 11.5% had 6-12 months, and 37.5% had over a year of experience.

**Secondary applicants.** A total of 127 secondary applicants (mean age 24.00 years, 60.63% female, and 93.70% White British) initially completed the secondary SJT at the interview stage of the selection process. 36 (24 female, 12 male) of these candidates completed either their PGCE (97.2%) or Bachelor's (2.8%) with the ITE provider. The mean age of secondary participants was 22.97 years ( $SD=2.76$ ), and 97.2% identified as being White British, and 2.8% Asian or Asian British. 91.7% of participants had prior experience working in schools, with 78.8% having less than one year of experience. 5.6% had less than 1 week of experience, 30.6% had 1-4 weeks, 16.7% had 1-3 months, 8.3% had 3-6 months, 11.1% had 6-12 months, and 19.4% had more than 1 year of experience.

## Procedure

**ITE provider's selection processes.** Applicants were initially selected to attend an interview based upon their academic merit (i.e. A-level scores or undergraduate degree transcripts) and the content of their UCAS personal statements. Applicants were also required to meet the threshold of the national professional maths (scored from 1-10) and English skills test (scored from 3 to 1) in order to receive an offer for the Bachelor or PGCE programs. Successful candidates were invited to attend an assessment day which included an online maths and written English test (combined duration 50 minutes), a group task (1 hour), a 10 minute interview with two interviewers (representatives from the university and partner schools), and a reflective task. Candidates are also required to complete a pre-interview essay (2-3 pages) regarding their assessment of the role of a teacher. Interviews took place on a rolling basis from November 2016 - July 2017. The tasks were designed by the ITE program staff to assess target attributes required for their course and the teaching profession, such as resilience, resourcefulness and motivation to teach. Candidates were assessed against a 4 point scale: areas of concern, good potential with support, good potential, and excellent potential. Applicants were required to obtain a score of 'good potential with support' in a set number of assessment categories in order to receive an offer. However, final selection decisions were based upon the assessments made by the two interviewers, and ratified by the head of the ITE program.

**Development of SJTs.** In the development of SJTs, researchers typically utilise an inductive critical incident approach to gather information from subject matter experts (SMEs) regarding challenging situations people in the field may encounter and the key attributes that are required for success in the relevant sector (Weekley, Ployhart, & Holtz, 2006). In contrast, deductive approaches select target attributes based on theories developed in previous research (Kasten & Freund, 2016; Motowidlo, Dunnette, & Carter, 1990). The 32 items included in each of the SJTs were initially developed using an inductive and deductive approach with the assistance of SMEs. To be deemed a SME, the person had to have prior teaching experience, interviewing experience in selection processes, or experience in observing novice teachers. Phase 1 was focused on identifying the target attributes of effective teachers and involved shadowing SMEs, conducting focus groups with stakeholders, domain rating questionnaires, and reviewing the target attributes identified. The central target domains identified were empathy and communication, organisation and planning, and resilience and adaptability. Phase 2 involved the development of item content, concordance panels, and the construction of pilot tests. As part of this stage, SMEs took part

in 45 minute interviews with the researchers, and were each asked to explain two complex situations that inexperienced teachers may face in the classroom, potential actions that may be taken in response to the given situation, and how they would rate each of the responses. The scenarios were then reviewed by the researchers to ensure each scenario was linked to the identified target attributes, before a panel of SMEs completed and commented on the test. The SMEs scoring of items was used as part of a consensus approach to develop the scoring key for the primary and secondary SJTs. Phase 3 of development comprised of piloting the SJTs with ITE applicants.

**Delivery of SJTs.** Applicants were initially invited to take part in the SJT in their interview invitation from the ITE provider, and completed the SJT online with an invigilator present following their interviews. The online SJT included a participant Information and consent statement that advised participants that participation was voluntary and they could withdraw from the study at any point. If participants agreed to the consent statement, they then proceeded to complete the SJT. The Primary and Secondary SJTs each contained 32 contextually rich scenarios, which were equally divided into two test groups (Group A and Group B) so that each participant would be presented with 16 scenarios which would take them approximately 30 minutes to complete. The tests contained three sections: (a) ranking items from most appropriate to least appropriate (b) picking the three most appropriate items, and (c) picking the least appropriate responses to the scenarios. An example scenario can be found in Figure 1.

**Primary SJT format and scoring.** For section A of the Primary SJT there were 10 (Group A) or 9 (Group B) scenarios with 5 item responses, which candidates had to rate from most (1) to least appropriate (5) in the given context, with 4 points available for each response item. The scoring for this section was based upon the ranking scoring system detailed by Patterson, Ashworth, and Good (2013) which allocates 4 points for correct responses, and deducts points based on the distance between the applicant's response and the correct answer (for example, if the correct answer was 'most appropriate' (1) and the candidate ranked the item as 2, they would receive 3 points as their answer was only one position away from the correct answer). For sections B and C, group A (6 questions) and group B (7 questions) were presented with scenarios and 8 item responses and were asked to select the three most appropriate options and one least appropriate option in the circumstances. 5 points were allocated for each correct response, equating to 20 available points per question (4 selection options x 5 maximum points). The total available score for group A and group B was 320 (16 scenarios x 20 maximum points).

**Secondary SJT format and scoring.** For the Secondary SJT, participants were also randomly allocated into two groups, completing 16 questions each. The same distance-measure scoring as described above was used. Section A 'rank items' contained 10 scenarios, and sections B 'pick three' and C 'pick least appropriate' contained 6 questions in both conditions. The total SJT available score for section A was 200 (10 scenarios, 5 response options x 4 maximum points), and 96 for sections B and C (6 scenarios x 16 maximum points for each question).

**Teaching placements.** Students who completed the Primary or Secondary PGCE were required to take part in two practical placements in schools over the course of the program. PGCE courses ran from September to July (38 weeks) and students were based in schools for 20 weeks. Placement 1 took place from October- January, with students initially based at their placement school for 3 days per week (for 4 weeks) and then 5 day per week thereafter. After 2-3 weeks at university and a one week placement in a different school setting, students started their second placement for 3 days per week for 4 weeks (February-March) increasing to 5 days per week up until the beginning of July. Students' performance over the course of each placement was assessed by their assigned mentor, and grades were

then moderated by the ITE provider. In line with higher education marking practices, students' placements were graded as follows: fail, 3 (lowest), 2, 2b, 2a, 1 (highest).

## Analysis

Analysis of the data included an analysis of descriptive statistics (mean, range, standard deviation) of the key variables. Correlation coefficients of SJT scores, interview scores, and teaching outcomes as measured by participants' placement grades were also conducted. Multiple linear regression was performed to show the contribution of SJT scores and interviews scores in predicting placement grades, and ANOVA was used to establish potential differences in placement grades for low and high performers on the SJTs.

## Results

Interview data and SJT scores collected as part of the initial ITE selection process were assessed for associations with student outcomes, as measured by placement grades for both primary and secondary candidates. Participant evaluations of the SJT were also examined.

### Primary

Table 1 shows the descriptive statistics of primary SJT scores, interview scores (including the maths and English application and test, group task, interview), and teaching placement grades (Placement 1 and 2). In Table 2, the results of bivariate correlations between SJT scores and teaching Placement 1 ( $r = .19, p = .07$ ) and Placement 2 ( $r = .24, p = .02$ ) are presented. Correlations with interview total scores ( $r = .15, p = .15$ ) and interview sub-scale scores did not reach significance. Interview total scores were not significantly related to Placement 1 grades ( $r = .09, p = .39$ ) or Placement 2 grades ( $r = .01, p = .92$ ). Placement 1 grades were associated with Placement 2 grades,  $r = .69, p < .001$ . Five questions were removed from the analysis as there was no consensus regarding the scoring key among SMEs. Table 3 shows the correlations between SJT scores with these five items removed and Placement 1 ( $r = .22, p = .03$ ) and Placement 2 grades ( $r = .30, p = .003$ ).

An independent samples t-test was conducted to compare SJT scores by gender. There was no significant difference in test scores between males ( $M = 195.79, SD = 22.41$ ) and females ( $M = 198.72, SD = 19.36$ ),  $t(94) = -1.4, p = .16$ . Bivariate correlations revealed that duration of prior experience in schools was not associated with participants' SJT scores ( $r = -.06, p = .57$ ), interview total scores ( $r = -.13, p = .21$ ), or Placement 1 ( $r = .19, p = .06$ ) or Placement 2 grades ( $r = .04, p = .69$ ).

Participants were divided into three equal groups based on their SJT scores (low, medium, and high scorers). A one-way ANOVA showed that there was a significant difference in Placement 2 grades between the 3 groups,  $F(2,93) = 5.16, p = .008$ . Post hoc comparisons using LSD test indicated that placement grades of high SJT scorers ( $M = 5.75, SD = .51$ ) were significantly higher than low scorers ( $M = 4.66, SD = 1.72, p = .003$ ) and medium scorers ( $M = 4.84, SD = 1.78, p = .02$ ). However, there was no significant difference in Placement 2 performance between low and medium scorers ( $p = .61$ ).

Multiple linear regression was conducted to test how primary SJT scores (with 5 items without scoring consensus removed) predicted Placement 2 grades after controlling for interview total scores. 9.1% of variance in Placement 2 grades was accounted for by interview and SJT scores,  $R^2 = .09, F(2,93) = 4.63, p = .01$ . SJT scores were found to be a significant predictor of Placement 2 grades ( $\beta = .31, t = 3.04, p = .003$ ), however interview scores were not found to be a statistically significant predictor of Placement 2 grades ( $\beta = -.04, t = -.38, p = .70$ ).

## Secondary

Table 4 shows the descriptive statistics Bivariate correlations were conducted to measure the relationship between SJT scores and Placement 1 ( $r = .09, p = .61$ ) and Placement 2 ( $r = .35, p < .05$ ). Correlations between interview scores and Placement 1 ( $r = .13, p = .44$ ) and Placement 2 ( $r = .03, p = .86$ ) did not reach statistical significance. Placement 1 grades were related to Placement 2 grades ( $r = .44, p = .007$ ). Length of prior experience in schools was significantly related to interview total scores ( $r = .46, p = .007$ ), and the assessment of subject knowledge during the interview process ( $r = .46, p = .008$ ). There was a negative correlation between duration of time spent gaining experience in schools prior to application and SJT scores ( $r = -.44, p = .01$ ) and Placement 2 grades ( $r = -.36, p = .03$ ).

An independent samples t-test revealed that there was no significant difference in SJT scores between females ( $M = 229.79, SD = 13.92$ ) and males ( $M = 228.33, SD = 11.77$ ),  $t(34) = -.31, p = .76$ . Candidates were split into 3 equal groups in terms of their performance on the SJT (low, medium, and high scorers). A one-way ANOVA revealed no significant difference in placement performance between the low ( $M = 3.75, SD = 2.14$ ) and high SJT scorers ( $M = 4.50, SD = 1.98$ ),  $F(2,33) = 2.14, p = .13$ . However, there was a significant difference between low and medium scorers ( $M = 5.33, SD = 1.44, p = .05$ ).

In a multiple linear regression model, 12.4% of variance in Placement 2 grades was accounted for by total interview scores and SJT scores, however this was not a statistically significant model,  $R^2 = .12, F(2, 33) = 2.34, p = .11$ . SJT scores were a significant predictor of Placement 2 grades ( $\beta = .05, t = 2.16, p = .04$ ), however interview total scores were not ( $\beta = .00, t = -.01, p = .99$ ). Linear regression with SJT scores only showed that SJT scores were a significant predictor of Placement 2 grades,  $R^2 = .12, F(1, 34) = 4.82, p = .04$ . Interview and SJT scores did not make a significant contribution in the prediction of Placement 1 grades,  $R^2 = .02, F(2, 33) = .35, p = .71$ .

## Participant evaluations of SJT

A random sample of candidates were selected to complete a feedback form following the completion of the SJT. The form reviewed applicants' evaluations of (a) relevance of the content, (b) level of difficulty, (c) fairness of content, (d) SJT differentiation, (e) fairness of SJT, (f) appropriateness of SJT, and (g) SJT measurement. Participants were required to rate each question from 1 = Strongly disagree to 5 = strongly agree. Table 6 shows primary and secondary candidates mean scores, with primary candidates' evaluation scores ranging from 3.52 to 4.42 and secondary applicants 3.33 to 4.36 out of 5.

Candidates were also asked to provide any additional comments on the SJT. The majority of participants provided positive feedback and suggested that the SJT was an effective measurement tool that presented challenging and realistic scenarios. For example, one participant noted 'I found the scenario based activities very effective, they get you to think about your actions and give you an insight into the many challenges and situations you will be part of when you are a teacher' (P34). Others noted that the SJT was different to other selection methods they had encountered, 'have not completed anything like it before. Was effective in trying to get you to think in a variety of ways' (P14). A small number of participants suggested that although they thought the SJT was useful, they thought it should be used in conjunction with other selection methods. Some candidates also shared their concerns that their inexperience as a teacher may have hindered their ability to answer some of the questions. For instance, one candidate noted 'I found the difficulty level appropriate and actually enjoyed trialling the test. However, I don't think this should be the only selection process or the majority of it as those having the tests will not know all the answers because they are not teachers yet, but are here to learn...' (P33).



## Discussion

The central aim of the study was to establish the relationship between ITE selection practices and candidates' subsequent performance on teaching placements, up to 18 months after selection. The results indicated that the primary SJT was related to candidates' Placement 1 and 2 grades, but more strongly correlated with their second (major) placement grades. The secondary SJT was related to candidates' Placement 2 grades, but not their Placement 1 grades. In contrast, the ITE providers' selection methods (interview, group task, maths and English assessment) were not associated with students' performance on either of the placements for primary or secondary applicants. There was no significant gender difference found between males and females in terms of SJT scores. The majority of candidates evaluated the SJT positively. The SJT may be a useful tool that could be deployed in combination with other ITE selection methods to provide objective information regarding the non-cognitive attributes of prospective teachers.

The results indicated that SJT scores, collected as part of the ITE selection process, were more strongly predictive of Placement 2 than Placement 1 grades. Considering Placement 2 took place towards the end of the ITE course, these findings may indicate that the predictive validity of the SJTs increases over time, which aligns with research in other fields (Buyse & Lievens, 2011; Blair, Hoffman, & Ladd, 2016; Lievens & Sackett, 2009). For example, in relation to dentistry selection processes, Buyse and Lievens (2011) established that the predictive validity of SJTs significantly increased from Year 1 to Year 5 of training, although the predictive validity of cognitive ability tests decreased over the same period. This may suggest that relationship between the targeted traits and behaviours assessed by SJTs and students' future performance may increase in significance over a period of time.

Although the positive correlation between SJT scores and student outcomes was moderate ( $r_s = .30$  to  $.36$ , roughly equivalent to Cohen's  $d$  of  $.60$  -  $.70$ ), the magnitude is similar to correlations found between SJTs and performance outcomes in other sectors (Christian, Edwards, & Bradley, 2010; McDaniel et al., 2011; Lievens et al., 2008), and it is considerably higher than the mean correlation of non-cognitive selection methods in education ( $r = .10$ ; Klassen & Kim, 2019). A meta-analysis conducted by McDaniel et al. (2011) in relation to job performance, established that the mean correlation across 95 studies was 0.34. In contrast, the ITE interview methods used in the current study (interview, group task, maths and English assessment) were not found to be associated with students' placement grades. This finding suggests that current ITE selection methods may not be objectively and systematically assessing attributes that may predict students' future performance on the course (Metzger & Wu, 2008; Pounder, 1989; Schumacher et al., 2013). Therefore, SJTs may assist in the selection process by providing an objective, evidence-based measure of candidates' non-cognitive attributes (Gold & Holodynski, 2015; Lievens et al., 2008). Utilising a combination of selection methods in the ITE selection process may provide a more in-depth assessment of candidates' attributes, and may increase the reliability and validity of selection decisions (Holden & Kitchen, 2016). This is arguably of central importance, considering that a key challenge facing the education sector is how to select and retain the most effective teachers that will have a positive impact on students' achievement and future prospects (Borman & Dowling, 2008; Rockoff et al., 2011).

In contrast to previous research that has established that females frequently perform better on SJTs in comparison to males by a mean of  $.10$  (O'Connell, Hartman, McDaniel, Grubb, & Lawrence, 2007), no significant gender differences were found in the current study in terms of SJT performance. This inconsistency in findings may be accountable to studies assessing SJTs that measure alternate constructs in differing contexts, and therefore the extent of gender differences may depend on the test's content (O'Connell et al., 2007; Whetzel, McDaniel, & Nguyen, 2008). This finding may warrant further investigation with a larger sample, however it may indicate that the content of the SJT does not adversely impact

candidates based on their gender. Developing an SJT that is fair all genders is arguably of importance, especially in the education sector, where less males apply to enter the profession (Cushman, 2007; McGrath & Van Bergen, 2017).

A further finding was that the majority of participants evaluated the SJT positively, and commented that it was an effective and 'different' selection tool that they enjoyed taking part in. Although the comments may have been influenced to some extent by responder bias, as candidates were asked to complete the feedback following the interview process and they may have wanted to emphasize positive aspects of the experience; this finding is consistent with several studies concerning candidates' reactions to SJTs (Chan & Schmitt, 1997; De Leng, Stegers-Jager, Born, & Themmen, 2018; Koczwara et al., 2012;). For example, Koczwara et al. (2012) assessed 260 junior doctors reactions to an SJT administered at a selection centre, and found that responses to the SJT were more favourable than some other measurement tools such as cognitive ability assessments. These findings may be due to the content of the SJTs being linked to their field of interest (De Leng et al., 2018). Candidates' evaluations of tests are important as this can influence their motivation to complete the test and their subsequent performance (Chan & Schmitt, 1997).

Considering the central aim of ITE selection processes is to select candidates who are likely to be successful on the program and in the teaching profession (Rockoff et al., 2011), the evaluation and identification of effective selection methods is of importance. Ensuring that the most effective teachers are employed has substantial socio-economic benefits, with evidence suggesting that quality teachers may increase students' earnings by 12% by the age of 28, and may improve higher education attendance rates among students (Chetty et al., 2014). The study has highlighted that SJTs may assist in the selection of effective teachers, as the attributes measured may be more strongly associated with students' practical performance in the field, in comparison to current selection tools. Traditional selection methods that evaluate students' abilities based on interviews have frequently be found to be limited in terms of their reliability, validity, and objectivity (McDaniels et al., 1994; Cook, 2016).

Although assessment centres have been found to have higher predictive validity (Hermelin, Lievens, Robertson, 2007; Patterson et al., 2016), they are often expensive to conduct and can only assess a small number of candidates at a time (Lievens et al., 2008). In contrast, once an SJT has been developed, it is comparatively cost effective to run, can be completed online with limited supervision, and has good validity in predicting candidates' future outcomes (Lievens et al., 2008; Whetzel & McDaniel, 2009;). Furthermore, the scoring of SJTs is developed by SMEs and applied consistently across candidates, potentially reducing the possibility of biases in selection decisions (Patterson et al., 2012). Therefore, SJTs may complement existing ITE selection practices by providing a cost-effective and objective method of measuring the non-cognitive attributes of prospective teachers.

### **Limitations**

Although the study has assessed the utility of SJTs in comparison to other ITE selection methods in terms of their ability to predict students' performance on teaching placements; the validity of the methods in terms of teaching performance once students have qualified is not known. Further research that measures the predictive validity of these measures over a longer period of time may be beneficial for establishing the effectiveness of ITE selection methods.

### **Conclusion**

In summary, the results from this research indicate that that an SJT may be useful in measuring prospective teachers' non-cognitive attributes at the application stage. In comparison to other sectors, little attention has been paid to evaluating current selection

---

tools and developing evidence-based methods that are reliable, valid, and fair. An SJT tailored for teacher selection purposes may provide ITE organisations with a simple and systematic way of evaluating large numbers of applicants, and may contribute to the selection of quality teachers who are dedicated to raising the achievement and future prospects of their students.

## References

- Ahmed, H., Rhydderch, M., & Matthews, P. (2012). Can knowledge tests and situational judgment tests predict selection centre performance?. *Medical Education*, 46(8), 777-784.
- Blair, C. A., Hoffman, B. J., & Ladd, R. T. (2016). Assessment centers vs situational judgment tests: longitudinal predictors of success. *Leadership & Organization Development Journal*, 37(7), 899-911.
- Borman, G. D., & Dowling, N. M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research*, 78(3), 367-409.
- Bowles, T., Hattie, J., Dinham, S., Scull, J., & Clinton, J. (2014). Proposing a comprehensive model for identifying teaching candidates. *Australian Educational Researcher*, 41(4), 365-380.
- Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, 31(4), 416-440.
- Byrnes, D. A., Kiger, G., & Shechtman, Z. (2003). Evaluating the use of group interviews to select students into teacher-education programs. *Journal of Teacher Education*, 54(2), 163-172.
- Buyse, T., & Lievens, F. (2011). Situational judgment tests as a new tool for dental student selection. *Journal of Dental Education*, 75(6), 743-749.
- Carter, A. (2015). Carter review of initial teacher training (ITT). Retrieved January 16, 2019, from [https://dera.ioe.ac.uk/21832/7/Carter\\_Review\\_16012015\\_Redacted.pdf](https://dera.ioe.ac.uk/21832/7/Carter_Review_16012015_Redacted.pdf)
- Chan D, & Schmitt N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143-159.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633-79.
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63(1), 83-117.
- Chung, R. R. (2008). Beyond assessment: Performance assessments in teacher education. *Teacher Education Quarterly*, 35(1), 7-28.
- Cohen, E., Hoz, R., & Kaplan, H. (2013). The practicum in preservice teacher education: a review of empirical studies. *Teaching Education*, 24, 345-380.
- Cook, M. (2016). *Personnel selection: Adding value through people - A changing picture*. Chichester, UK: John Wiley & Sons.
- Cushman, P. (2007). The male teacher shortage: A synthesis of research and worldwide strategies for addressing the shortage. *KEDI Journal of Educational Policy*, 4(1), 79-98.
- Denner, P. R., Salzman, S. A., & Newsome, J. D. (2001). Selecting the qualified: A standards-based teacher education admission process. *Journal of Personnel Evaluation in Education*, 15(3), 165-180.
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237-251.
- Gold, B., & Holodyski, M. (2015). Development and construct validation of a situational judgment test of strategic knowledge of classroom management in elementary schools. *Educational Assessment*, 20(3), 226-248.
- Goldhaber, D., Grout, C., & Huntington-Klein, N. (2015). Screen Twice, Cut Once: Assessing the Predictive Validity of Teacher Selection Tools. *Society for Research on Educational Effectiveness*.
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100(2), 267-71.
- Hanushek, E. A., & Zhang, L. (2009). Quality-consistent estimates of international schooling and skill gradients. *Journal of Human Capital*, 3(2), 107-143.

- Hobson, A. J. et al. (2010), *International Approaches to Teacher Selection and Recruitment*, OECD Education Working Papers, No. 47, OECD Publishing.  
<http://dx.doi.org/10.1787/5kmbpnhh6qmx-en>
- Holden, M., & Kitchen, J. (2016). Choosing how we choose: A review of teacher education admissions literature. *Teaching and Learning*, 11(1), 16-31.
- Kasten, N., & Freund, P. A. (2015). A meta-analytical multilevel reliability generalization of Situational Judgment Tests (SJTs). *European Journal of Psychological Assessment*, 32(3), 230-240.
- Klassen, R., & Kim, L. (2018, November 10). *Development of an Online Construct-Informed Situational Judgment Test for Screening Applicants for Initial Teacher Education*. TSP Working Paper, <https://doi.org/10.31234/osf.io/64hry>
- Klassen, R. M., & Kim, L. E. (2019). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review*, 26, 32-51.
- Koczwara, A., Patterson, F., Zibarras, L., Kerrin, M., Irish, B., & Wilkinson, M. (2012). Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Medical Education*, 46(4), 399-408.
- Lievens, F., Patterson, F., Corstjens, J., Martin, S., & Nicholson, S. (2016). Widening access in selection using situational judgement tests: evidence from the UKCAT. *Medical Education*, 50(6), 624-636.
- Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, 37(4), 426-441.
- Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, 97(2), 460-468.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79(4), 599.
- McGrath, K. F., & Van Bergen, P. (2017). Are male teachers headed for extinction? The 50-year decline of male teachers in Australia. *Economics of Education Review*, 60, 159-167.
- Metzger, S. A., & Wu, M. J. (2008). Commercial teacher selection instruments: The validity of selecting teachers through beliefs, attitudes, and values. *Review of Educational Research*, 78(4), 921-940.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75(6), 640-647.
- Murnane, R. J., Willett, J. B., Duhaldeborde, Y., & Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management*, 19(4), 547-568.
- National Council for Accreditation of Teacher Education. (2010). *Transforming teacher education through clinical practice: a national strategy to prepare effective teachers*. Retrieved from <https://eric.ed.gov/?id=ED512807>
- O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb III, W. L., & Lawrence, A. (2007). Incremental validity of situational judgment tests for task and contextual job performance. *International Journal of Selection and Assessment*, 15(1), 19-29.
- Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M., & O'Neill, P. (2012). Evaluations of situational judgement tests to assess non-academic attributes in selection. *Medical Education*, 46(9), 850-868.
- Patterson, F., Ashworth, V., & Good, D. (2013). Situational judgement tests: A guide for applicants to the UK Foundation Programme. *Medical Schools Council*, 1-29.
- Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education*, 50(1), 36-60.
- Pounder, D. G. (1989). Improving the predictive validity of teacher selection decisions: Lessons from teacher appraisal. *Journal of Personnel Evaluation in Education*, 2(2), 141-150.
- Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, 6(1), 43-74.

- Salvatori, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. *Advances in Health Sciences Education*, 6(2), 159-175.
- Schumacher, G., Grigsby, B., & Vesey, W. (2015). Determining effective teaching behaviors through the hiring process. *International Journal of Educational Management*, 29(1), 139-155.
- Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. *Situational judgment tests: Theory, measurement, and application*, 26, 157-182.
- Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, 19(3), 188-202.
- Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, 21(3), 291-309.

**Table 1***Descriptive Statistics for Primary SJT, Interview Scores, and Placement Scores (n=96)*

	<i>M (Range)</i>	<i>SD</i>
SJT score	239.57 (185-278)	15.55
SJT score (with 5 items removed)	197.86 (156-242)	20.22
Maths Audit (professional skills)	6.60 (2-10)	1.77
English Audit (professional skills)	1.9 (1-3)	.42
Interview: Maths application and test	4.60 (1-8)	1.56
Interview: English application and test	4.74 (1-8)	1.36
Interview: communication in presentation	3.04 (2-4)	.65
Interview: awareness of the everyday	2.98 (1-4)	.78
Interview: motivation to teach	3.10 (2-4)	.73
Interview: communication of subject	2.89 (2-4)	.71
Interview: target qualities	3.01 (2-4)	.75
Total interview score	32.81 (22-44)	5.31
Placement 1	4.69 (1-6)	1.15
Placement 2	5.08 (1-6)	1.52

*Maths Audit is scored 1(lowest) to 10 (highest). English Audit is scored 1(highest) to 3 (lowest).*

**Table 2***Correlations Between Primary SJT Score, Interview Scores, and Placement Scores (N=96)*

	1	2	3	4	5	6	7
1. SJT score	-	.15	.03	.11	.15	.19	.24*
2. Maths and English Application		-	.57**	.04	.70**	.07	.003
3. Maths and English Audit (professional skills)			-	.21*	.74**	-.11	-.13
4. Interview: Teacher qualities total scores				-	.69**	.16	.09
5. Total Interview Scores					-	.09	.01
6. Teaching Placement 1						-	.69**
7. Teaching Placement 2							-

\* $p < .05$ , \*\* $p < .001$



**Table 3***Correlations Between Primary SJT Score (With 5 Scenarios Removed), Interview Scores, and Placement Scores (N=96)*

	1	2	3	4	5	6	7
1. SJT score	-	.09	.04	.17	.16	.22*	.30**
2. Interview: Maths and English Application and test scores		-	.57**	.04	.70**	.07	.003
3. Maths and English Audit (professional skills)			-	.21*	.74**	-.11	-.13
4. Interview: Teacher qualities total scores				-	.69**	.16	.09
5. Total Interview Scores					-	.09	.01
6. Teaching Placement 1						-	.69**
7. Teaching Placement 2							-

\*p&lt;.05, p&lt;.001

**Table 4***Descriptive Statistics for Secondary SJT, Interview Scores, and Placement Scores*

	<i>N</i>	<i>M (Range)</i>	<i>SD</i>
SJT	36	229.31 (184-252)	13.09
Group task 1 (Interpersonal)	36	3.39 (2-4)	.60
Group task 2 (Leadership)	36	3.28 (2-4)	.66
Interview: Teacher qualities	36	3.36 (2-4)	.64
Interview: Wider awareness	35	3.22 (2-4)	.90
Interview: Resilience	36	3.08 (2-4)	.60
Interview: Subject knowledge	35	3.23 (2-4)	.73
Total interview score	36	19.39 (12-24)	3.14
Placement 1	36	3.42 (1-6)	1.48
Placement 2	36	4.52 (1-6)	1.93

**Table 5***Correlations Between Secondary SJT Score, Interview Scores, and Placement Scores (N=36)*

	1	2	3	4	5	6	7	8	9	10
1. SJT score	-	-.06	.003	.05	.04	.12	.06	.04	.09	.35*
2. Interview: Group task 1		-	.51**	.52**	.41*	.46**	.39*	.66**	.04	-.08
3. Interview: Group task 2			-	.50**	.48**	.30	.34*	.72**	-.09	-.14
4. Interview: Teacher qualities				-	.63**	.59**	.56**	.83**	.11	.07
5. Interview: Wider awareness					-	.55**	.61**	.81**	.29	.15
6. Interview: Resilience						-	.35*	.68**	.06	.01
7. Interview: Subject knowledge							-	.73**	.34*	.25
8. Total Interview Scores								-	.12	.02
9. Teaching Placement 1									-	.44**
10. Teaching Placement 2										-

\*p&lt;.05, \*\*p&lt;.001

**Table 6**

*Primary Applicants' Evaluation Responses for the Situational Judgment Test (1= Strongly Disagree, 5= Strongly Agree)*

Item content	Primary			Secondary			MIN	MAX
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Q1. Overall, the content of the SJT was clearly relevant to those applying	173	4.42	0.84	42	4.33	1.00	1	5
Q2. Overall, the level of difficulty of the SJT was appropriate for those applying	173	4.24	0.79	42	4.33	0.79	1	5
Q3. Overall, the content of the SJT appeared to be fair to those applying	173	4.41	0.81	42	4.36	0.91	1	5
Q4. The SJT will help to differentiate between candidates applying	172	3.79	1.00	42	3.52	1.17	1	5
Q5. The SJT is a fair method of selection as part of the Selection Process	170	3.52	1.07	40	3.33	1.14	1	5
Q6. The SJT is an appropriate method as part of the Selection Process	170	3.58	0.96	40	3.53	1.01	1	5
Q7. The SJT is able to measure the attributes that are necessary for Newly Qualified Teachers	170	3.92	0.98	40	3.75	1.10	1	5

Figure 1. Example of rank item question from situational judgment test.

You have explained a writing task and answered all of the questions that students have asked. As the students begin the task, one student, Ella, starts to throw pens around and is distracting the other students in the class. You know from previous incidents that Ella often becomes frustrated when she does not understand how to complete activities and that she often displays her frustration by being disruptive. What should you do?

*Rank in order the following actions in response to this situation (1= Most appropriate; 5= Least appropriate).*

- A. Provide Ella with a textbook, that will give more information and enable her to complete the exercise
- B. Encourage Ella, by telling her that she is capable of completing the activity
- C. Send Ella out the class if she continues to be disruptive
- D. Tell Ella that if she continues to throw pens then you will give her a lunchtime detention
- E. Give Ella an easier task to complete