**Developing a Situational Judgment Test for Admission into Initial Teacher Education in Oman: An Exploratory Study**

**Developing a Situational Judgement Test for Admission into Initial Teacher Education in Oman: An Exploratory Study**

# Abstract

Selecting applicants into initial teacher education programs (ITEPs) in Oman focuses mainly on their academic achievement in school. Applicants’ non-cognitive attributes (NCAs), such as resilience and motivation, receive less attention in the acceptance decision. Situational Judgment tests (SJTs) are kind of simulation tests which have been used in the selection process of other professions with promising results. Here, we aim to (a) develop an SJT for selecting applicants into ITEPs in Oman, (b) explore its’ initial properties in terms of reliability and validity, and (c) explore applicants’ reactions to the developed test. The research design consists of four phases. In Phase 1, we utilize a multi-step design to explore the target NCAs underlying teacher effectiveness in Oman. The non-cognitive attributes were used in Phase 2 to guide development of an SJT with groups of working teachers (N = 116). The developed SJT was initially piloted in Phase 3 with new entrants to an ITEP (N = 171) and then piloted with other criterion measures in Phase 4 (N = 142). Results from Phase 4 reveal that the developed SJT has good internal consistency (α = .75). Scores on the SJT correlate significantly with scores on two facets of the Big-five personality measure (i.e. conscientiousness and agreeableness). In addition, the SJT has a medium positive significant correlation with the participants’ GPA, and a negative non-significant correlation with the interview scores. The participants’ responses to the SJT were positive. With further research, we recommended SJTs as promising measurement for selecting prospective teachers in Oman.

***Key words:***teacher selection; initial teacher education; non-cognitive attributes; situational judgment tests.

# Introduction

Admissions practices for initial teacher education programmes (ITEPs), including Oman, tend to focus on applicants’ cognitive characteristics, looking at performance on examinations or admissions tests that focus on subject knowledge or reasoning ability (Ingvarson et al., 2013; Casey & Childs, 2007). Applicants’ non-cognitive attributes (NCAs) - such as interpersonal skills, motivation, and resilience - receive less attention in the admission process. However, studies in ITEPs (e.g., Dolan, 2012; Fantilli & McDougall, 2009; Jacobowitz, 1994) point to the urgent need to evaluate applicants’ NCAs during the ITEP admission process in a way that is reliable, valid, fair, and efficient. In addition, the limitations of current selection measures (Author(s), 2017b) indicate that ITEPs need better selection measures to identify the best possible candidates. Research from other professional fields (e.g., medicine) point to the use of situational judgment tests (SJTs) to select applicants, with data suggesting that the tests have good levels of reliability, predictive validity, and incremental validity for testing non-cognitive attributes related to professional practice (Patterson et al., 2012a). An SJT is a simulation test where the applicant is presented with a variety of situations he would be likely to meet on the job and asked to identify the appropriate response(s) (Weekley & Ployhart, 2006; McDaniel & Nguyen, 2001). Fig. 1 presents an example of an item used in an SJT to that tests a teacher’s judgement in classroom interactions.

***Figure 1 about here***

Some initial research on using SJTs for the selection of prospective teachers has been conducted (e.g., Author(s), 2014, 2017a), with promising results, encouraging further research in different contexts. The current study extends the initial work conducted in the UK and elsewhere to the context of Oman and aims to (a) develop an SJT for selecting applicants to the ITEPs in Oman and (b) explore its reliability and validity and applicants’ reactions to the test. To our knowledge, this the first study for developing SJTs for selecting prospective teacher in Oman. The results are intended to enhance the selection practices and put the floor for further studies.

# The Problem in Oman

Education in Oman is growing rapidly. Starting with only three schools in 1970, the country has worked to expand its education to all. In 2015, the gross enrolment ratio (GER) was approximately 100% for students in grades 1-10 and 95% for grades 11 and 12. The mean class size was approximately 26 students per class and the student-teacher ratio was 9.5:1 (OEC, 2016). Despite the increase in educational provision, the main challenge to the basic education system in Oman is low student achievement in comparison with the achievements of students internationally (Oman newspaper, 2014). One of the causes of this challenge is believed to be the relatively low quality of teachers in the country (OMoE, 2012).

Although there is no single definition of teacher effectiveness, studies show that NCAs such as interpersonal relationships and personality traits are important domains for teacher effectiveness in Oman (Al Barwani et al., 2012; Al-Ani et al., 2012). In addition, NCAs such as commitment, motivation, and attitudes towards teaching are believed to be a significant matter for teachers in Oman. Due to the importance of these relationships, researchers recommend modifying the selection criteria for trainee teachers to ensure a better selection of suitable candidates (Issan, 2011; Chapman et al., 2012; Al Barwani, 2002; Al Harthy et al., 2013).

The policy of ITEPs in Oman has been developed to ensure a high quality of teacher preparation, but the selection process has remained unchanged, and is largely based on academic attributes with little concern for other applicant characteristics. Although the ITEP entrance requirements attract students with strong academic performance, a study investigating the career paths of ITEP graduates showed that half of the participants are only somewhat or not at all committed to teaching as a career at the time of graduation (Chapman et al., 2012). Surprisingly, research indicates a nonsignificant relation between secondary school certificate results and student teacher performance at the ITEP (Al Barwani, 2002).

# Non-cognitive Attributes for Selection into ITEP

Research on teachers’ NCAs identifies them as potential predictors of student achievement and teaching quality. Teachers’ enthusiasm, for example, positively predicts student interest (Keller et al., 2014) and has a positive effect on pupil motivation (Kunter et al., 2013a). Furthermore, teacher-student relations have a significant impact on teachers’ job satisfaction (OECD, 2014). However, research into NCAs reveals less consensus on their definition and identification. NCAs are referenced in the literature using different terminology, such as ‘soft skills’, ‘personality traits’, ‘non-academic abilities’, and ‘character’. Here, we use ‘NCAs’ to refer to within-person variables, which might include beliefs, motives, personality traits, and dispositions (Author(s), 2017a).

The model developed by Kunter et al. (2013b) in Fig. 2 highlights the role of NCAs in the professional development of prospective teachers. In the model, the process of developing prospective teachers’ competence and behaviours through learning opportunities is influenced by two main factors: contextual aspects (education policy, characteristics of the ITEP, and so on), and personal characteristics (cognitive and non-cognitive attributes, and background factors). The model illustrates the importance of considering applicants’ personal characteristics at the selection process for an ITEP.

***Figure 2 about here***

Sautelle et al. (2015) identified six psychological constructs of effective teaching that can be assessed when selecting teachers for training programmes and which are likely to allow differentiation between candidates; namely, extraversion, agreeableness, conscientiousness, resilience, self- regulation, and cognitive ability. This work is partly based on the Big Five personality model (McCrae & Costa, 2004; Rammstedt & John, 2007), with research showing a relation between the Big Five dimensions and job performance (Rothmann & Coetzer; 2003; Barrick & Mount; 1991).

However, there is no single NCA that can be named the sole predictor of future success (McGeown et al., 2015). Hence, different attributes related to teacher effectiveness should be measured during the ITEP selection process. Casey and Childs (2007) state that it is important to distinguish between attributes that can and cannot be learned in a teacher education programme. They claim that, during the selection process, successful applicants will show elements of the necessary attributes that are not easily learned in the programme.

Another challenge identified in the literature is finding the appropriate method of measurement for the targeted attribute(s). Wang and Fwu (2007) note that most ITEPs use more than one means of evaluating a single criterion. Despite the variety in selection measures, research on their influence on teacher effectiveness is scarce. In their meta-analysis, Authors(s) (2017b) found a small effect size for the selection methods for both cognitive and non-cognitive admission tests. In addition, Casey and Childs (2011) noted no significant relationship between two admission criteria and the readiness of teacher candidates in mathematics. Caskey et al. (2001) stated that the search for the most effective admission selection procedures should continue.

# Situational Judgment Tests (SJTs)

Unlike personality tests, SJTs are designed to evaluate individuals’ judgement in contextualised workplace settings (Ryan & Ployhart, 2014). The tests are currently used to screen applicants to medical school in the UK, Belgium, Canada, and Australia (Patterson et al., 2015). The development process for the SJT begins with the collection of critical incidents and response options, targeting a set of competencies and attributes from the job analysis. The answer key (the responses’ effectiveness) and the scoring method are then developed by a group of experts. Participants in the development process are mostly experienced professionals and/or subject matter experts (Weekley, Ployhart & Holtz, 2006).

Although some studies claim that SJTs can measure individual or specific groups of constructs, SJTs are more effective as a measure of multiple constructs and hard to be isolated from general cognitive ability. Hence, the research indicates that it is better to view them as a method that can be designed to measure a variety of cognitive and non-cognitive constructs, rather than a single construct (Whetzel & McDaniel, 2009; Lievens, 2006).

The work of Motowidlo and his colleagues (Motowidlo et al., 1990; Motowidlo et al., 2006) provides a better understanding of the theoretical basis of the SJTs. In explanation, the researchers offer the ‘implicit trait policy’ theory, which suggests differences between individuals in their implicit beliefs about the importance of personality traits for determining behavioural effectiveness. In other words, the effectiveness judgment for a situation can be related to the individual’s weight to the trait in that specific situation, which reflects fundamental socialization processes (parents, schooling, etc.) and personal dispositions (Lievens & Motowidlo, 2016; Whetzel & McDaniel, 2009). Therefore, understanding applicants’ implicit attributes can help to predict their future effectiveness.

The reliability of SJTs has been analysed in reviews and meta-analyses. One meta-analysis by McDaniel et al. (2001) found Cronbach’s alpha coefficients ranging from 0.43 to 0.94. Patterson et al. (2015) state that the internal consistency of the SJTs used in medical and dental contexts is approximately *α* = 0.7 or more. Generally, and regardless of which test is used, the research broadly indicates that SJTs have moderate to good levels of reliability (Patterson et al., 2015).

Looking at the construct validity of the tests, Christian et al. (2010) found that the mean validity of the studies on SJTs’ measurement of teamwork skills was .38, leadership skills was .28, interpersonal skills was .25, and conscientiousness was .24. In Patterson et al. (2016), the predictive validity of an SJT for entry into postgraduate general practice (GP) specialty training in Australia was found significantly predicted all three end-of-training assessments (*r* = .12 to .54). Furthermore, a systemic review for research in the selection for medical school, Patterson et al. (2012a) show that the SJTs are the best single predictor of performance, with an incremental predictive power over cognitively oriented tests.

Finally, candidate reactions to the content and the use of SJTs in the admission process for medical schools are often found to be positive (Patterson et al., 2015). Similarly, Author (2014) found that SJTs were favoured by most of the applicants to initial teacher training programmes in the UK. The positive reactions to the selection measure could have significant consequences: increasing acceptance of the selection outputs and decreasing the possibility of claims (Weekley & Ployhart, 2006).

# Aim and Research Questions

The current study starts by exploring the key NCAs required for effective teaching in Oman; aiming (a) to develop an SJT for measuring the NCAs of applicants to the ITEPs in Oman, (b) to explore the initial findings on the reliability and validity of the test; and (c) to note the applicants’ reactions to the SJT.

The study aims to address the following research questions:

1. What are the key NCAs considered necessary for prospective teachers in the government schools (grades 5-12) in Oman, as identified from official documents and from stakeholders’ perspectives?
2. To what extent can the SJTs be used in the ITEP admission process in Oman to better understand the NCAs of the new undergraduate applicants? That is,

* What is the reliability (the internal consistency) of the developed SJTs in Oman?
* What is the criterion-related validity of the developed SJTs in Oman? That is, how do the SJT scores correlate with three criterion measures: the applicants’ scores in the admission interview, students’ academic performance (GPA), and the Big Five inventory (BFI)?
* What are the applicants’ reactions to the SJT’s content and use as a selection tool?

# Method and Results

Following theoretical and practical approaches for developing SJT (Patterson et al., 2016; Weekley, Ployhart, & Holtz, 2006), and building on the initial work in the UK (Author(s), 2017a), the method in this study includes eleven steps divided into four phases (see Fig. 3). Participants were stakeholders from the Ministry of Education (MoE) and ITEPs in Oman (college tutors, schools’ principals, supervisors, teachers in the government school (grades 5-12), and new students on the ITEP). We focused on school teachers (grades 5-12) because most ITEP graduates in Oman are recruited to teach these grades. Official permission was obtained from both the MoE and two ITEPs in Oman, allowing the first Author to collect data from the targeted samples. At all phases, participants were asked to express their agreement to participate by reading and signing a detailed consent form.

***Figure 3 about here***

## Phase One (steps 1-4): Identifying the Target Attributes (SJTs’ specifications)

The goal in this phase was to identify the key attributes related to teacher effectiveness in Omani schools (grades 5-12), for which applicants to the ITEP should be screened. These attributes form the inputs for the development of the SJT items in Phase two. An explorative research approach was taken, comprising four steps as follow.

In step one, official documents from the MoE and an ITEP in Oman were reviewed. Then, in step two, an explorative semi-structured interview was conducted with key stakeholders (college tutors on an ITEP (*n* = 2), school principals (*n* = 3), and teachers’ supervisors (*n* = 3). The content analysis of the related official documents and the exploratory interviews produced a list of 56 attributes. Similar attributes were aggregated, which produced a second list of 43 attributes. .

In step three, the 43 attributes were put into a matrix including the attributes of the three domains identified in the UK namely, ‘empathy and communication, organization and planning, and resilience and adaptability’. This led to the establishment of two new domains in the context of Oman: enthusiasm and motivation, and professional ethics. In addition, and due to the discussions with interviewees, the domain of ‘empathy and communication’ was changed to ‘communication skills’.

In step four, the importance of the five domains was explored by using a self-constructed questionnaire with a 10-point scale (1 = not important, 10 = very important) and a sample of teachers, supervisors, and school principals from three educational governorates in Oman (*N* = 181). Of the respondents, 58% were female and the mean work experience duration was 16.4 years. Data from the questionnaire showed that the five domains were considered to be important for teachers in Oman, with a mean ranging from 9.21 for ‘resilience and adaptability’ (*SD* = 1.60) to 9.53 for ‘professional ethics’ (*SD* = 1.54). The mean was also high for the importance of the domains to the ITEP applicants, ranging from 8.92 for ‘resilience and adaptability’ (*SD* = 1.71) to 9.24 for ‘enthusiasm and motivation’ (*SD* = 1.65).

The completion of this Phase resulted in a framework, with five key non-cognitive domains and 29 attributes: ‘communication skills’ (seven attributes), organisation and planning’ (five attributes), ‘resilience and adaptability’ (six attributes), ‘enthusiasm and motivation’ (five attributes) and ‘professional ethics’ (six attributes). The five non-cognitive domains and their definitions formed the basis for developing the SJTs for Oman, and presented in Figure 4.

***Figure 4 about here***

## Phase Two (steps 5-9): Developing the SJT for Oman

The process of developing the SJT consisted of five steps (as shown in Fig. 3). The participants were working teachers from a number of government schools (grades 5-12) in an educational governorate in Oman. They were senior teachers or recommended by their school principals as good teachers.

In step five, the aim was to collect scenarios and responses (i.e. SJTs items) targeting the five domains found in Phase one. The SJT items were collected from two sources. Firstly, the initial SJT built in the UK was translated from English into Arabic by a bilingual translator, producing 34 items. Then, a sample of working teachers in Oman were presented by examples of the translated items and asked to write new items. In this way, 88 items (i.e. scenarios and responses) were generated. The relation between the items and the targeted domains was investigated in step seven.

In step six, the88 items were revised by the first author and 12 were removed due to unclarity in the content or similarity with other item(s). Following the format used in the SJT developed in the UK, the remaining 67 items were written in two formats: ‘ranking’ and ‘select best three’. The ‘ranking’ format had 19 items with five responses for each. The test-taker was asked to read the scenario in each item and rank the responses from the most appropriate (ranked as 1) to the least appropriate (ranked as 5).While the ‘select best three’ format had 48 items with six option for each item and the test-taker was asked to select the most appropriate three options to the given scenario in each item.

In step seven, the 67-item SJT was presented to a group of expert teachers (*N* = 8) from four different schools, each of whom had been recommended as a good teacher by the school principal. The teachers revised each item in terms of clarity and suitability for the Omani context and for new entrants into ITEPs. Data analysed qualitatively and indicated a closed agreement between teachers in the suitability of 53 items. Thirteen items were removed due to the teachers’ feedback (i.e. not suitable for new entrants, not appropriate for the Omani context). Participants were also asked to identify which of the five domains can be measured by each of the 53 items. Teachers related 30 items to a specific one domain. They believed that 14 items represented ‘communication skills’, four items for ‘resilience and adaptability’, two items for ‘organisation and planning’, three items for ‘enthusiasm and motivation’, and seven items for ‘professional ethics’. For the remaining 23 items, teachers assumed that each item could explain two or more domains (e.g. an item can explain both ‘communication skills’ and ‘professional ethics’, another item can be related to three domains together) Therefore, the SJTs items did not map clearly onto the five domains. This multidimensional nature of SJT (i.e. one item can target more than one domain) is not unique for this study but supported by similar findings in the literature (Kasten & Freund, 2015; Schmitt & Chan, 2006).

As there is often no single correct answer with an SJT, the answer key refers to the best judgment of a situation from a set of given responses (i.e., what is the right answer most likely to be?). In step eight, we develop an answer key for the 53-item SJT by using the rational approach (i.e. consensus in answers) (Weekley et al., 2006). A group of teachers (*N* = 108) from ten schools, who had not participated in the previous steps, were asked to participate. Of these, 48.1% were female and the mean work experience duration was 12.9 years. Using the rational model of seeking consensus, teachers were asked the following: ‘*Consider yourself as a novice teacher, what should you do in the following situations (rank/ select best three)’*. Teachers’ answers in each item were aggregated and put into percentage for each option. Data indicated that there was strong consensus in the answers for 38 items (14 in ranking and 24 in select best three). However, the data for the remaining 15 items showed poor agreement. For instance, in some items there was a big difference in the selected options between male teachers and the females. Other items had also high missing data (left unanswered). Therefore, we decided to remove those 15 items and did not include them in the next phase.

In step nine, we develop the scoring key used and implemented in selection for medical schools (Metcalfe & Dev, 2013) and in the work of Author(s) (2017a). For the ‘ranking’ items, the responses were scored according to their distance to the answer deemed correct, while in the ‘select best three’ items, participant was awarded four points for each correct option and zero points for selecting the wrong option – no negative marking was used.

## Phase Three (step 10): Piloting the SJT

Here, the 38-item SJT built in Phase two was piloted with its scoring key. The aim was to explore the quality of the items before tested with other criterion measures in the next phase.

Participants. The participants were a convenience sample of first-year students on an ITEP in Oman (*N* = 171; 53.4% female). The participants’ subjects were biology (33.3%), chemistry (22.2%), physics (19.4%), maths (23.6%) and English (0.7%).

Measure. The 38-item SJT developed in Phase two was used. There were 24 items in the ‘select best three’ format, with six options for each item, and 14 items in the ‘ranking’ format, with five options for each item. It was presented in Arabic and in a paper-based format. The test ended with an open-ended question *(‘Please give your view on the test in terms of its suitability for future use in the admission procedure for the ITEP, given the need to ensure the selection of the best possible future teachers. Please also share any other comments you have on the test’)*. The total scores were 568 points for the 38 items.

Procedure. The pilot was conducted during the induction week at the beginning of the academic year 2016/2017, in one ITEP in Oman. The participants were invited to take the test as a one group at a predetermined time. The participants were given an open-ended amount of time to finish the test, and most participants did so in approximately 40-50 minutes.

Participants’ answers to the SJT were scored and analysed by using SPSS 25. Descriptive statistics showed that data was close to normal distribution and females scored significantly better than males (*t* (112) = 7.65, *p* < .001). the SJT had a good internal consistency (*α* = .81 for the 38 items).

The quality of the 38 items was investigated by using some statistical techniques used in similar studies (Klassen et al., 2017b; Patterson et al., 2015b). In each type of questions, we analysed each item by looking to the item difficulty, difference in item difficulty by gender, item-total correlation, and Cronbach’s alpha if the item is deleted. In addition to the inclusion of the statistical findings, the decision about including or excluding each item was also based on the content and domain the item purported to measure.. This process further reduced the number of items to 29. Responses to the open-ended question include good feedback about the test and its suitability for use in the selection process.

## Phase Four (step 11): Testing the SJT (reliability, validity, and applicants’ reaction)

The aim of this phase was to explore the properties of the 29-item SJT found in Phase three by looking to the reliability (internal consistency), validity (the correlation between the SJT scores and other criterion measures), and the applicants’ reactions to the content and suitability of the test.

**Participants.** The participants were a convenience sample of students at an ITEP in Oman (*N* = 142, 74% female). The majority of the participants (87%) were in their first or second year. However, 18 participants were in their third, fourth, or fifth year. The data of those 18 students were excluded due to their involvement in teaching practices courses. Almost all the subjects were represented in the sample, with a majority studying English (33%).

**Measures.** To explore the validity of the SJT, three criterion measures were used: personality (using the Big Five Inventory, BFI); cognitive abilities, using the students’ latest GPA; and participants’ performance in current selection measure, using the interview scores. Besides, a feedback form was used to evaluate the applicants’ reaction. All measures are described as follow:

***SJT.*** The 29-item SJT developed in Phase three was used. It was presented in Arabic and in a pencil and paper format. The total score of the SJT was 460 points.

***Personality.*** To explore the construct validity of the SJT, participants were asked to complete an Arabic version of the 60-item BFI tested with a sample of graduated students in Oman by Kazem (2002), measuring: neuroticism, extroversion, openness, agreeableness, and conscientiousness. These five dimensions of personality share common attributes with the five domains shaped the developed SJTs in Oman. The BFI consists of 60 items: 12 items for each factor with some reverse items. Participants scored each item on a five-point scale as follows: (1) strongly disagree, (2) disagree, (3) neutral, (4) agree, and (5) strongly agree. The total possible score was 60 points for each factor..

***Current admission tool.*** The interview scores for the participants at admission were obtained from the ITEP. in order to obtain the concurrent validity of the developed SJT (Patterson et al., 2012a).The interview in the ITEP admission process in Oman is conducted to evaluate the applicants on eight items: showing care for academic specialisation; interested in teaching all categories of students, including special education needs; demonstration of appreciation for Islamic and Omani values; showing care for scientific research; demonstrating problem-solving and decision-making skills; good knowledge of the role of technology in education; ability to communicate verbally in an effective manner; and possessing charisma and demonstrating a professional appearance. The maximum score for the interview was 24.

***Cognitive ability.*** Although the developed SJTs target non-cognitive attributes, related systemic review and meta-analysis studies found that SJTs overlap with cognitive abilities (Patterson et al., 2012a). Thus, the correlation between the SJT and student cognitive abilities in term of GPA was tested in this study. GPA is a measure of students’ academic achievement in ITEPs in Oman and ranges between ‘0’ to ‘4’. It has been used in different studies in Oman to explore the influence of academic and non-academic factors on students’ achievement (Alkhausi et al., 2015

***Applicants’ reaction.*** Applicant reactions were measured using an Arabic translation of the measure developed in the UK by Author(s) (2014b). It had seven items asking participants to rate the SJT in terms of relevance, difficulty, fairness, differentiation, and appropriateness on a 5-point scale (1 = strongly disagree, 5 = strongly agree). In addition, participants were given the opportunity to provide comments about the SJT in response to an open-ended question.

**Procedure.** Data collection took place at an ITEP in Oman (different from the one in Phase three). The maximum time allowed was one hour, but most students finished all of the measures within 40 minutes. The participants provided their university number for the three measures, to match their scores with their marks during the admission interview and their cumulative GPA.

*Descriptive statistics:* An analysis of the data from Phase four (*N* = 124) shows that the total scores ranged from 250 to 404 (*M* = 360; *SD* = 26.6). The mean makes about 78% of the total possible scores (total possible scores were 460) which indicates the difficulty of the test. The difficulty of the ‘select best three’ items was approximately 77%, and 79% for the ‘ranking’. The scores were significantly higher for females than for males (female: *M* = 365, *SD* = 21.9; male: *M* = 339, *SD* = 32.9; *t* (30) = 3.77, *p* = .001). The data were slightly negatively skewed, indicating that most participants scored high in the test.

*Reliability:* the reliability of the SJT was tested in terms of its internal consistency, using Cronbach’s coefficient alpha. The results show that the internal consistency of the SJT was *α* = .75 for the 29 items: .80 for the ‘select best three’ items and .55 for the ‘ranking’ items.

*Validity:* the correlation between the SJT and the other measures is shown in Table (1). Firstly, the data show that correlation was positively small and significant with conscientiousness (*r* = 0.29, *p* = .002, *n* = 111) and agreeableness (*r* = 0.19, *p* = .04, *n* = 109). It was also positive with extraversion (*r* = 0.11, *p* = .26, *n* = 109) and openness (*r* = 0.18, *p* = .06, *n* = 108), though not statistically significant. SJT had a negative non-significant correlation with neuroticism (*r* = -0.16, *p* = .10, *n* = 104). Secondly, there was a negative non-significant correlation between scores in the SJT and participants’ performance at interview (*r* = -0.17, *p* = .11, *n* = 89). Finally, a positive moderate and significant correlation was found between the SJT and participant GPA (*r* = .31, *p* = .001, *n* = 121).

***Table 1 about here***

*Applicants’ reaction:* a total of 140 participants completed the applicant feedback measure, whilst 67 participants answered the open-ended question. First, the responses to the seven statements were analysed in terms of means and standard deviations, as shown in Table (2). The data indicate a strong agreement with all statements, the mean ranging from 3.82 to 4.50, in the scale of 1-5 options.

***Table 2 about here***

More than 91% of the participants agreed or strongly agreed that the content of the SJT was relevant to teaching. The percentage who ‘agreed’ or ‘strongly agreed’ that the contents were of the ‘appropriate level of difficulty’ was approximately 82%, and approximately 75% agreed that it was ‘fair’. In contrast, 77% and 68% agreed or strongly agreed, respectively, that the SJT was ‘able to differentiate between the candidates’ and ‘able to measure NCAs’, whilst approximately the same percentage (61%) agreed on the ‘fairness’ and ‘appropriateness’ of the SJT as a selection method.

Finally, about 73% of the participants’ comments about the SJT in response to the open-ended question were positive. Some of the comments sate that SJT in general: ‘good test’, ‘excellent’, ‘very useful’, ‘clarifies the nature of teaching as a profession’, and ‘makes me more interested in teaching’. Furthermore, some comments supported the appropriateness of the SJT for selecting candidates onto the ITEP: ‘must be implemented’, ‘better than the interview’, ‘better than just seeing results in secondary school’, and ‘could be included alongside the current tools’. However, few comments challenged the suitability of the SJT for use as a selection method, and the clarity of some of the items.

# Discussion

Here, the findings are discussed and interpreted according to the aims of the study.

## Key NCAs of Prospective Teachers in Oman

We develop the SJT in Oman by firstly exploring the key NCAs that seen necessary for screening applicants into ITEPs. These attributes shaped the development of the SJTs items in Phase two. By involving key stakeholders (i.e. ITEPs’ tutors, school principals, supervisors, working teachers), results revealed a list of 29 attributes divided into five domains: ‘communication skills’ (seven attributes), organisation and planning’ (five attributes), ‘resilience and adaptability’ (six attributes), ‘enthusiasm and motivation’ (five attributes) and ‘professional ethics’ (six attributes). In a 10-point scale, participants rated highly the important to consider these domains in the selection process (8.92 ≤ *M* ≤ 9.24) to ensure the effectiveness of prospective teachers.

The similarities and differences between the three domains found in the UK study and the five domains used in the current study offer a novel understanding of the influence of the national and cultural contexts on perceptions of the critical NCAs of effective teachers (Author(s), 2018). In addition, our findings align with the national educational policies and current practices in selecting and evaluating teachers in Oman.

The first domain, ‘communication skills’, influences teacher and student outcomes (OECD, 2014). In Oman, the job description of a teacher clearly emphasises domain, and teachers are also annually evaluated on their relationships with the school, peers, pupils, and parents. The second domain, ‘resilience and adaptability’, has been shown to be related to teacher effectiveness (Gu and Day, 2007). The attributes related to this domain are clearly observed in the official documents relating to teaching in Oman. For example, teachers are required to accept advice and feedback and to show confidence. The third domain, ‘organisation and planning’ domain – is one of the important dimensions in the Pianta and Hamre CLASS framework (2009) of teacher effectiveness. The planning dimension is also required for teachers in Oman. Teachers are asked in their job description to prepare and present an annual timeline for their duties, and they are then evaluated according to their submitted plans.

The fourth domain, ‘enthusiasm and motivation’ domain is strongly associated with teacher effectiveness in Oman. Several studies have shown that motivation is a significant matter for current working teachers and has a negative impact on students’ achievement in college (Al Harthy et al., 2013; Chapman et al., 2012). The emphasis of this domain in the Omani context reflects the need for enthusiastic and motivated teachers. Hobson et al. (2009) stated that a prior commitment to the profession is very important, and Keller et al. (2014) reveal that teacher enthusiasm positively predicts students’ interest.

The fifth domain, ‘professional ethics’ has been used here as a translation of the Arabic ‘Akhlaqyaat Almihnah’, which is widely used in the education system in Oman. It appears in the first statement of the teacher job description. Campbell (2003) suggested that ethics lay at the heart of teacher professionalism, and Lumpkin (2008) stated that teachers must be viewed as morally sound role models for young people to follow, trustworthy, honest, and respectful.

To conclude, the findings from a multi-step process in Phase one introduced a better understanding to the key attributes that related to teacher effectiveness in Oman and their important to be used in screening applicants into ITEPs. Although the current interview process used for admission ITEPs in Oman include questions about at least three of the five domains (enthusiasm and motivation, professional ethics, and communication skills), the outcomes of the interview process have little influence on the decision regarding acceptance onto the ITEP. This asserts the need to enhance the selection method to give the same attention to non-cognitive attributes as given to cognitive abilities.

## Exploring the Initial Properties of the Developed SJT

The study evaluated the initial proprieties of the test in terms of: (a) its reliability (internal consistency), (b) validity (correlation with the criterion measures), and (c) the applicants’ reactions to the content and the proposed use of the test as a selection measure.Our findings showed good reliability for the developed SJT, and some evidence of construct and concurrent validity was found.

***Correlation with the Big Five.*** Our results indicate a small correlation of between .1 and .3 for the participants' SJT scores and BFI scores. The significant correlation with conscientiousness and agreeableness might be explained by looking to their components and how they relate to the five domains shaping the SJTs. Conscientiousness, in the BFI, contains items related to self-control, active process of planning, organising, hardworking, responsibility, and the application of moral principles. Thus, it shares some of the attributes of at least three of the five domains: ‘enthusiasm and motivation’, ‘planning and organisation’, and ‘professional ethics’. Agreeableness, however, includes sympathy with others, eagerness to help, and unselfishness. Thus, it shares attributes with ‘communication skills’ and ‘professional ethics’. This assumption of the similarity between the attributes measured by the SJTs and the BFI factors needs further research. This result aligns with the findings of the meta-analysis by McDaniel et al. (2007) which found similar correlations for SJTs and Big Five traits. In Chan and Schmitt (2002), SJTs were also found to be significantly, albeit weakly, correlated with big five personality traits (r ranged from .19 to .29).

***Correlation with interview.*** Although the developed SJT and the interview both test the NCAs of the participants, the results illustrate a negative non-significant correlation between the scores (*r* = -0.17, *p* = .11). Our findings contradict those of Author(s) (2017a) in the UK which found a correlation of .29 with the overall interview score. A correlation of *r* = 0.52 was also found between applicant scores on the SJT and the multiple-mini interviews during the selection process in medical schools (Patterson et al., 2012a). The limitation in the current interview process might has an effect in the negative correlation with the SJT. Although the interview process for an ITEP in Oman targets the personality of the applicant, its validity has not – to our knowledge – previously been measured. We suggest that further research address this limitation and conduct the study during the admission process.

***Correlation with GPA.*** The results show a medium positive significant correlation between the SJT and participant GPA (*r* = .31, *p* = .001). This aligns with previous studies on the relationship between NCAs, measured by the SJT, and students’ academic achievement, measured by GPA. In Lievens (2013), a video-based SJT measuring interpersonal behaviour had significant added value over cognitive tests for predicting interpersonal GPA and doctor performance. Furthermore, in some related systemic reviews and meta-analyses, the SJT has a correlation with cognitive abilities (Patterson et al., 2012a). This correlation was found to be 0.46 in the McDaniel et al. (2001) meta-analysis. However, working with a sample of employees, Chan and Schmitt (2002) found SJTs to be uncorrelated with cognitive ability (*r* = –.02). This variability in findings was examined by McDaniel et al. (2001), who found that SJTs based on job analyses and those with detailed questions were more highly correlated with cognitive ability.

**Applicants' reactions.** The content was declared relevant, of an appropriate difficulty, and fair most candidates. Specifically, the majority agreed or strongly agreed that the content of the SJT was relevant and fair for those applying for the ITEP and that the level of difficulty was appropriate. In the open-ended question, the majority of the comments were positive about the content of the SJTs. In general, the applicants’ reactions to the developed SJT – as content and as a selection tool – were positive. This positive reaction was also seen in the studies conducted in the UK (Author(s), 2014) and Australia (other & co-author, 2017).

# Limitations

Despite the promising findings, the study has limitations. The participants were a non-random sample of teachers, recruited from just one educational governorate. A further study should use a random sample from different regions and include other stakeholders to build the SJT. In addition, as it was not possible to conduct the pilot study at the selection stage, the participants in Phase four were not entirely fresh entrants. This overlap in participants may had influenced the correlation between the SJT and the other criterion measures (such as interview scores). In addition, the moderate and weak correlation suggest replicating the results with further studies with a bigger sample in a high-stake context. Finally, and to best predict the validity of the SJT, the current results should be correlated with measures related to teacher effectiveness. This was not possible in this study, but it is recommended for future research.

# Conclusion

The findings from study contribute to an understanding of the educational context in Oman by showing how a new test of NCAs could enhance the current selection practices used for ITEP admissions. It reveals the importance of measuring the non-cognitive attributes of prospective teachers as early as possible; during the entrance to the ITEP. As a recommended solution, the SJT developed in this study could be considered an additional and promising selection measurement, benefiting from further studies and evaluation.

# References

Author(s) (2018).

Author(s) (2017a).

Author(s) (2014).

Author(s) (2017b).

Author(s) (2016).

Other & co-author (2017).

Al Barwani, H. D. T. A. (2002). *Bridging the gap between secondary education, higher education, and the world of work*. In International Conference on the refrom of secondary education, 22-24/12/2002, Muscat.

Al Barwani, T. A., Al-Ani, W., & Amzat, I. (2012). An effective teaching model for public school teachers in the Sultanate of Oman: Students' stance. *Education, Business and Society: Contemporary Middle Eastern Issues*, *5*(1), 23-46.

Al Harthy, S. S. H., Jamaluddin, S., & Abedalaziz, N. A. (2013). Teachers’ attitudes and performance: An analysis of effects due to teaching experience. *International Interdisciplinary Journal of Education*, *2*(9), 889-894.

Al-Ani, W. T., Al-Barwani, T., & Al-Buloshi, S. (2012). Defining effective teaching in Oman: Searching for consensus? *Educational Research, 3*(1), 87-96.

Al-Harthy, I. S., & Aldhafri, S. S. (2014). The relationship among task-value, self-efficacy and academic achievement in Omani students at Sultan Qaboos University. *International Review of Social Sciences and Humanities*, *7*(2), 15-22

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta‐analysis. *Personnel psychology*, *44*(1), 1-26.)

Campbell, E. (2003). Moral lessons: The ethical role of teachers. *Educational Research and Evaluation*, *9*(1), 25-50.

Casey, C., & Childs, R. (2007). Teacher education program admission criteria and what beginning teachers need to know to be successful teachers. *Canadian Journal of Educational Administration and Policy*, (67)

Casey, C., & Childs, R. (2011). Teacher education admission criteria as measure of preparedness for teaching. *Canadian Journal of Education*, *34*(2),

Caskey, M. M., Peterson, K. D., & Temple, J. B. (2001). Complex admission selection procedures for a graduate preservice teacher education program. *Teacher Education Quarterly*, 7-21.

Chamorro‐Premuzic, T., & Furnham, A. (2004). A possible model for understanding the personality‐intelligence interface. *British Journal of Psychology*, *95*(2), 249-264

Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, *15*(3), 233-254.

Chapman, D. W., Al-Barwani, T., Al Mawali, F., & Green, E. (2012). Ambivalent journey: Teacher career paths in Oman. *International Review of Education*, *58*(3), 387-403.

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

Dolan, A. M. (2012). Reforming teacher education in the context of lifelong learning: the case of the BEd degree programme in Ireland. *European Journal of Teacher Education*, *35*(4), 463-479.

Fantilli, R. D., & McDougall, D. E. (2009). A study of novice teachers: Challenges and supports in the first years. *Teaching and teacher education*, *25*(6), 814-825.

Gu, Q., & Day, C. (2007). Teachers resilience: A necessary condition for effectiveness. *Teaching and Teacher education*, *23*(8), 1302-1316

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics, 19*(4), 451-464.

Hobson, A. J., Giannakaki, M.-S., & Chambers, G. N. (2009). Who withdraws from initial teacher preparation programmes and why? *Educational Research, 51*(3), 321-340.

Ingvarson, L., Schwille, J., Tatto, M. T., Rowley, G., Peck, R., & Senk, S. L. (2013). *An Analysis of Teacher Education Context, Structure, and Quality-Assurance Arrangements in TEDS-M Countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)*: ERIC.

Issan, S. A., Al-Nabhani, H. Z., Kazem, A. M., & Al-Ani, W. T. (2011). Omani teachers’ attitudes towards teaching as a profession. *Indian Journal of psychology & Education*, *1*(1), 25-40.

Jacobowitz, T. (1994). Admission to teacher education programs: Goodlad's sixth postulate. *Journal of Teacher Education, 45*(1), 46-52.

Kazem, A.(2002). Psychological values and the big five factors of personality. *Journal of Educational and Psychological Studies, 3 (1),*11-42, (in Arabic).

Keller, M. M., Goetz, T., Becker, E. S., Morger, V., & Hensley, L. (2014). Feeling and showing: A new conceptualization of dispositional teacher enthusiasm and its relation to students' interest. *Learning and Instruction, 33*, 29-38.

Kunter, M., Kleickmann, T., Klusmann, U. & Richter, D. (2013a). The development of teachers’ professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Eds.) *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp.63–77). New York: Springer.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013b). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105(3)*, 805.

Lievens, F. (2006). International situational judgment tests. *Situational judgment tests: Theory, measurement, and application*, 279-300.

Lievens, F. (2013). Adjusting medical school admission: assessing interpersonal skills using situational judgement tests. *Medical Education*, *47*(2), 182-189.:

Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology*, 1-20.

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, 37(4), 426-441.

Lumpkin, A. (2008). Teachers as role models teaching character and moral virtues. *Journal of Physical Education, Recreation & Dance*, *79*(2), 45-50.

Metcalfe, D., & Dev, H. (2013). *Oxford Assess and Progress: Situational Judgement Test*. Oxford University Press.

McAbee, S. T., & Oswald, F. L. (2013). The criterion-related validity of personality measures for predicting GPA: A meta-analytic validity competition. *Psychological Assessment*, *25*(2), 532.

McCrae, R. R., & Costa Jr, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and individual differences*, *36*(3), 587-596

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*(1/2), 103-113.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. (2007). Situational judgment tests, response instructions, and validity: a meta‐analysis. *Personnel Psychology, 60*(1), 63-91.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: a clarification of the literature. *Journal of Applied Psychology, 86*(4), 730

McGeown, S. P., St Clair-Thompson, H., & Clough, P. (2015). The study of non-cognitive attributes in education: proposing the mental toughness framework. *Educational Review*, 1-18.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*, 640–647.

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). A theoretical basis for situational judgment tests. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 57-81)

OECD. (2014). *TALIS 2013 Results: An International Perspective on Teaching and Learning*: OECD Publishing.

Oman Education Council (OEC). (2016). *The annual report about Education in Oman*. Muscat (Arabic)

Oman Ministry of Education (OMoE). 2012*. Education in Oman: The Drive for Quality: Jointly prepared by The Ministry of Education and The World Bank*. Muscat

Oman newspaper. (2014), Press report of a Seminar in Human resources and empowerment of the future, in <https://educouncil.gov.om> (in Arabic)

Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M., & O’Neill, P. (2012a). Evaluations of situational judgement tests to assess non‐academic attributes in selection. *Medical education, 46*(9), 850-868.

Patterson, F., Lievens, F., Kerrin, M., Zibarras, L., & Carette, B. (2012b). Designing Selection Systems for Medicine: The importance of balancing predictive and political validity in high‐stakes selection contexts. *International Journal of Selection and Assessment, 20*(4), 486-496.

Patterson, F., Zibarras, L., & Ashworth, V. (2015). Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100. *Med Teach*, 1-15.

Patterson, F., Rowett, E., Hale, R., Grant, M., Roberts, C., Cousans, F., & Martin, S. (2016). The predictive validity of a situational judgement test and multiple-mini interview for entry into postgraduate training in Australia. *BMC medical education*, *16*(1), 87.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational researcher*, *38*(2), 109-119.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality*, *41*(1), 203-212

Rothmann, S., & Coetzer, E. P. (2003). The big five personality dimensions and job performance. *SA Journal of Industrial Psychology*, *29*(1), 68-74.

Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual review of psychology*, *65*, 693-717.

Sautelle, E., Bowles, T., Hattie, J., & Arifin, D. N. (2015). Personality, resilience, self-regulation and cognitive ability relevant to teacher selection. *Australian Journal of Teacher Education*, *40*(4),

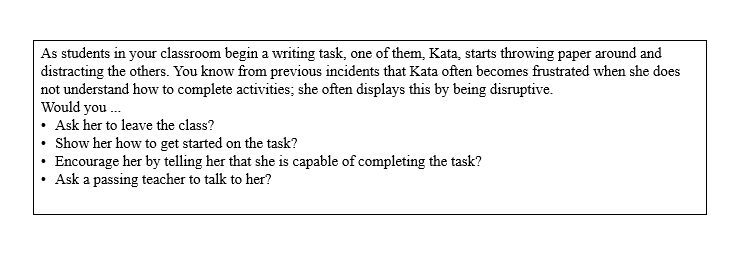
Schmitt, N., & Chan, D. (2006). Situational judgment tests: Method or construct? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 135-155). New Jersey

Wang, H. H., & Fwu, B. J. (2007). In pursuit of teacher quality in diversity: A study of the selection mechanisms of new secondary teacher education programmes in Taiwan. *International Journal of Educational Development*, *27*(2), 166-181.

Weekley, J. A., & Ployhart, R. E. (2006). *Situational judgment tests: Theory, measurement, and application*: Psychology Press.

Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J. A. Weekley & R. E. ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 157-182).

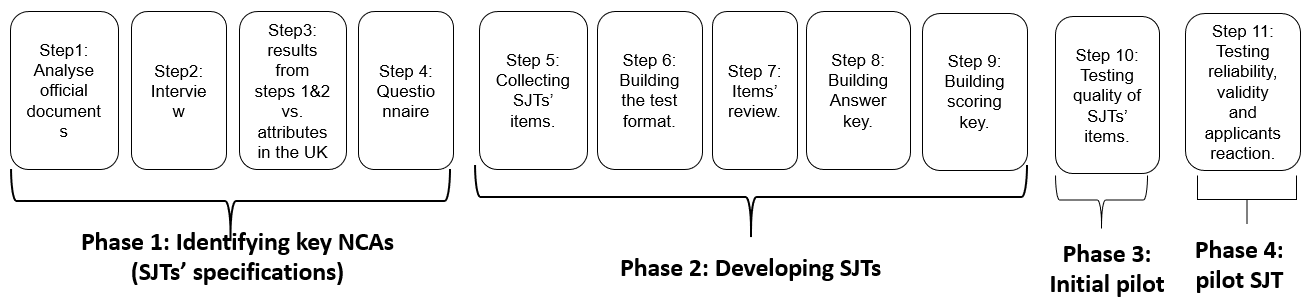
Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review, 19*(3), 188-202.



*Figure 1.*Example of an item used in SJTs (From Author(s), 2016)



*Figure 2.*Model of the development of teacher effectiveness (in Author(s), 2017; adapted from Kunter et al., 2013).



*Figure 3.* A summary of the four phases of the study

****

*Figure 4.* Definitions of the five non-cognitive domains found in Oman

Table 1

*SJTs’ Correlation with the Other Measures*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Neuroticism | Extraversion | Openness | Agreeableness | Conscientiousness | Interview scores | GPA |
| SJTs Total scores | Pearson Correlation | -.16 | .11 | .18 | .19\* | .29\*\* | -.17 | .31\*\* |
| Sig. (2-tailed) | .10 | .26 | .06 | .04 | .002 | .11 | .001 |
| N | 104 | 109 | 108 | 109 | 111 | 89 | 121 |
| \*. Correlation is significant at the 0.05 level (2-tailed). | | | | | | | | | |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). | | | | | | | | | |

Table 2

*Mean and standard deviation of Participants’ Feedback on the SJTs*

|  |  |  |  |
| --- | --- | --- | --- |
|  | N | Mean | Std. Deviation |
| 1. Overall, the content of the SJT was clearly relevant to those applying for the ITEPs to be teachers in the future | 140 | 4.50 | .835 |
| 1. Overall, the level of difficulty of the SJT was appropriate for those applying for the ITEPs to be teachers in the future | 140 | 4.15 | .848 |
| 1. Overall, the content of the SJT appeared to be fair for those applying for the ITEPs to be teachers in the future | 140 | 4.00 | .890 |
| 1. The SJT will help to differentiate between candidates applying for the ITEPs to be teachers in the future | 140 | 3.99 | .944 |
| 1. The SJT is a fair method that can be used as part of the selection process for candidates applying for the ITEPs to be teachers in the future | 140 | 3.74 | 1.01 |
| 1. The SJT is an appropriate method that can be used as part of the selection process for candidates applying for the ITEPs to be teachers in the future | 139 | 3.82 | .972 |
| 1. The SJT is able to measure the NCAs that are necessary for teachers | 140 | 3.86 | .918 |